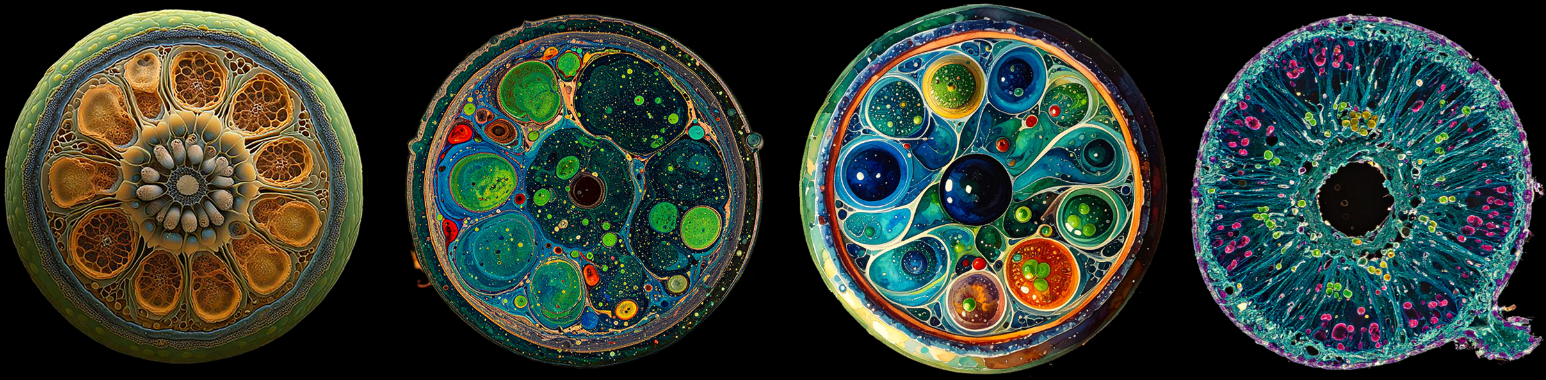
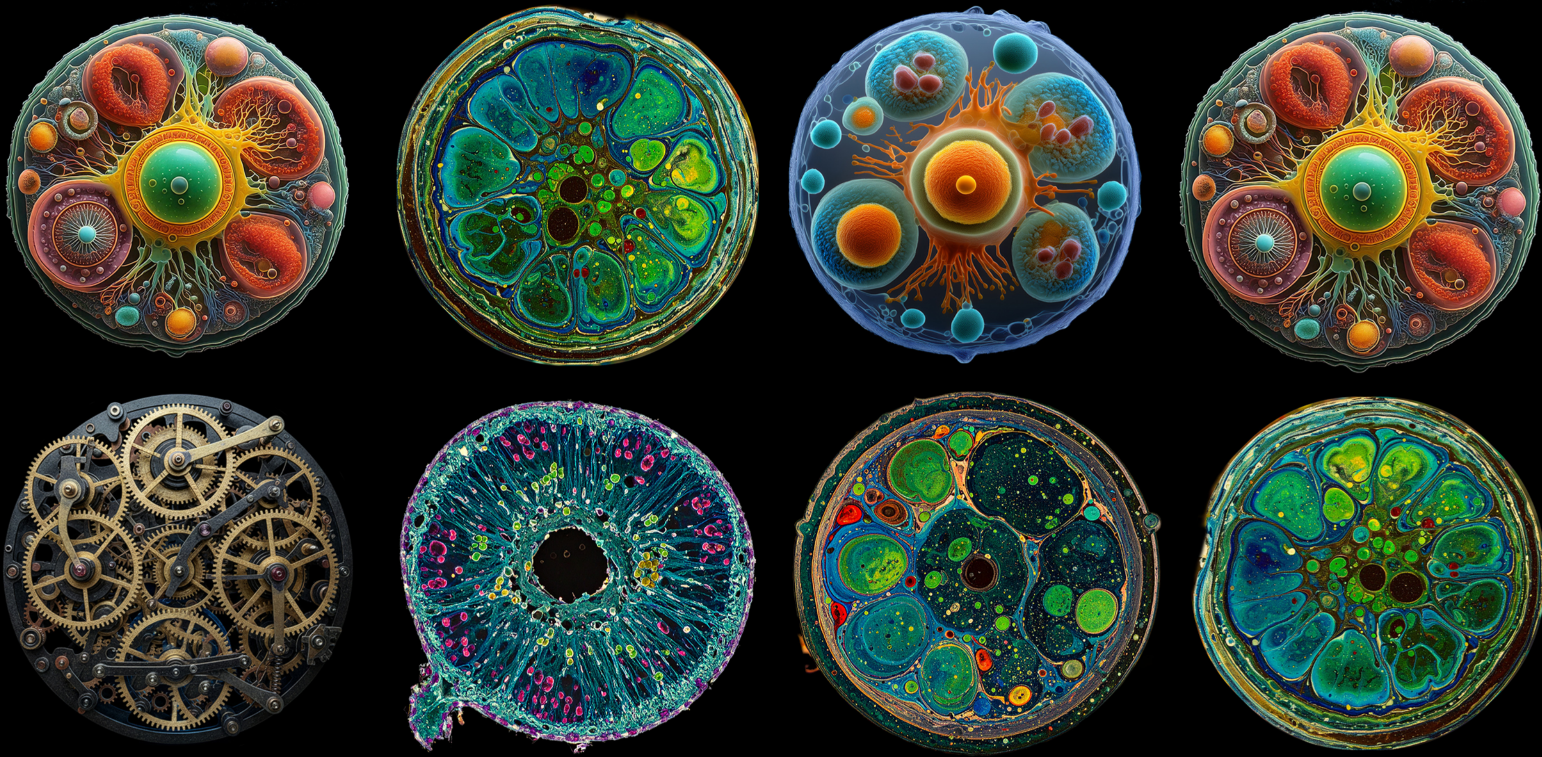
Organisms

Journal of Biological Sciences



What AI Can Learn from Biology



Guest Editor Ali Hossaini

Vol. 8, No. 1-2 (2025)

ISSN: 2532-5876

SAPIENZA
UNIVERSITÀ DI ROMA

Organisms

EDITORS

Guest Editor

Ali Hossaini, Department of Engineering, King's College London (United Kingdom)

Editor-in-Chief

Andràs Paldi, École Pratique des Hautes Études, Paris (France)

Associate Editors

Mariano Bizzarri (former Editor-in-Chief), Department of Experimental Medicine & Systems Biology Group, Sapienza University of Rome (Italy)

Giuseppe Longo, Centre Cavaillès, ENS, Paris (France)

Ana Soto, Tufts University School of Medicine, Boston (USA)

Carlos Sonnenschein, Tufts University School of Medicine, Boston (USA)

Editorial Board

Sergio Adamo, Department of Histology & Medical Embriology Sapienza University of Rome (Italy); **Stuart G. Baker**, Division of Cancer Prevention, National Cancer Institute, Rockville (USA); **Alessandra Cucina**, Sapienza University of Rome (Italy); **Barbara Demeneix**, Museum Nationale d'Histoire Naturelle, Paris (France); **Evelyn Fox-Keller**, History and Philosophy of Science Massachusetts Institute of Technology, Boston (USA); **Scott F. Gilbert**, Department of Biology, Swarthmore College, Philadelphia, PA (USA); **Alessandro Giuliani**, Istituto Superiore di Sanità, Rome (Italy); **Rodolfo Guzzi**, Optical Society of America (Italy); **Sui Huang**, Institute for Systems Biology, Seattle, WA (USA); **Paul-Antoine Miquel**, Laboratoire Erraphis Université de Toulouse 2 (France); **Vidyanand Nanjundiah**, Department of Molecular Reproduction, Development and Genetics and Centre for Ecological Sciences, Indian Institute of Science, Bangalore (India); **Nadine Peyriéras**, Institut de Neurobiologie Alfred Fessard, CNRS, Gif-sur-Yvette (France); **Olaf Wolkenhauer**, Department of Biology and Bioinformatics, University of Rostock (Germany)

Managing Editor

Stefano Serafini, Systems Biology Group, Sapienza University of Rome (Italy)

ORGANISMS is a biannual, open access, peer-reviewed, online journal devoted to basic, translational, theoretical and clinical research. Topics of interest include but are not limited to the following areas: complexity in biology; carcinogenesis; theory of biological organization; biophysics; mathematical and computational approaches in biology and medicine; developmental processes, and degenerative diseases.

ORGANISMS' Editorial Board is involved in research programs into many aspects of normal and pathological inflammation. The Board promises to conduct and facilitate unbiased and critical reviews of manuscripts. Every proposal submitted for publication is read at least by an editor for a preliminary review. If the paper matches the editorial policies and quality level, then it is sent to reviewers for a blind evaluation via the OJS platform.

ORGANISMS provides immediate open access to its content on the principle that making research freely available to the public supports a greater global exchange of knowledge. Authors of articles published in ORGANISMS retain the copyright of their works. Authors can archive pre-print, post-print, and publisher's versions.

The journal is committed to maintaining the highest level of integrity in the content published. All research must have been carried out within an appropriate ethical framework. If there is suspicion that work has not taken place within an appropriate ethical framework, Editors will follow the COPE guidelines and may reject the manuscript, and/or contact the author(s)' institution or ethics committee.

Research involving human subjects, human material, or human data, must have been performed in accordance with the Declaration of Helsinki, and must have been approved by an appropriate ethics committee. A statement detailing this, including the name of the ethics committee and the reference number where appropriate, must appear in all manuscripts reporting such research. If a study has been granted an exemption from requiring ethics approval, this should also be detailed in the manuscript (including the name of the ethics committee that granted the exemption). Further information and documentation to support this should be made available to the Editors on request. Manuscripts may be rejected if the Editor considers that the research has not been carried out within an appropriate ethical framework. Editors may contact the ethics committee for further information.

When reporting experiments on animals, authors should indicate whether the institutional and national guide for the care and use of laboratory animals was followed. See International Committee of Medical Journal Editors ("Uniform Requirements for Manuscripts Submitted to Biomedical Journals"), February 2006.

https://rosa.uniroma1.it/rosa04/organisms/index

CONTENTS

Special Issue
“What AI Can Learn from Biology”

Editorial

Forthcoming in January 2026

Cover image: *Organism/Mechanism*, Ali Hossaini, 2025

Organisms

SAPIENZA
UNIVERSITÀ DI ROMA

Editorial

Vol. 8, No. 1–2 (2025)
ISSN: 2532-5876

DOI: 10.13133/2532-5876/19209

Special Issue, “What AI Can Learn from Biology”

Ali Hossaini[a*]

[a] *Department of Engineering, King’s College London, United Kingdom*

***Corresponding author:** Ali Hossaini, Email: ali.hossaini@kcl.ac.uk

Abstract

This special issue uses AI to cast light on the nature of life. Many assume that life emerges from a blend of information and complexity. If this is the case, then we might expect a future generation of machines to exhibit lifelike behavior or, as some would claim, to come alive. Two perspectives are offered for considering the question of life: agency and intelligence. Intelligence is associated with information, rationality and consequent knowledge representations, while agency associates with embodiment, judgment and material organization. Predictions about machine life rely on conceptions of intelligence, but the addition of agency to the analysis of life and lifelike behavior results in nuanced conclusions that can beneficially inform regulation and future research.

Keywords: AI, cognition, worlding, thermodynamics

Citation: Hossaini, A, 2025 “Special Issue, ‘What AI Can Learn from Biology’” [editorial], *Organisms: Journal of Biological Sciences*, vol. 8, no. 1–2, pp. 5–6. DOI: 10.13133/2532-5876/19209

Humanity has never been so successful, or so threatened, during its short time on earth. This special issue of *Organisms* addresses two challenges of the polycrisis described by Greg Anderson in “AI in This World and the Next”: biology and intelligence. Contributors to this issue view these challenges through approaches which outrun dominant paradigms about information, emergence and complexity. One of the great conundrums of biology is its subject matter. Unlike other disciplines, notably physics and chemistry, the subjects of biology – organisms – self-organize, replicate, evolve and proliferate. The bright line between organisms and abiotic phenomena, including mechanisms, is the Second Law of Thermodynamics. Put simply but accurately, life does not run down, and any explanation of life must account for its persistence across four billion years. Though rarely addressed, the relevance of biology to AI is clear. Cognition evolved as a function of life, and, starting with imprints on bone, wood and clay, cognitive prostheses have accounted for humanity’s early evolutionary success and our growing potential for failure. Humanity has a long history of engineering, and inventions which enhance our faculties have inspired awe since the first cities. Whether we take the perspective of life or history, AI is a step change not a revolution.

The concept for this special issue began in 2016 at the London-based think tank RUSI. I was asked by Randolph Kent, then a Fellow, to consider whether AI might become hostile. That inquiry resulted in the article, “Modelling the Threat from AI: Putting Agency on the Agenda”, and responses from four biologists who informed my position. Denis Noble and Ray Noble emphasize the production of novelty by organisms, while Ana Soto and Carlos Sonnenschein emphasize the materiality of life. *RUSI Journal* published the articles in 2019 (in an issue also devoted to AI), and they are included here. Since then, I have continued to contemplate agency and intelligence as competing paradigms for our self-understanding. A happier world could keep these questions in the philosopher’s den, but, in the present age, we project our self-understanding haphazardly and invisibly into engineering projects of increasing scale. By clarifying what intelligence is and contrasting it with agency, we also clarify our role – and that of engineering – in the life world.

We have decided to release the contents of this issue in themed instalments. Greg Anderson’s aforementioned essay opens the section on agency, and the issue as a whole, with the argument that human conceptions are fundamentally fluid. In contrast to other contributors, Anderson is an historian. I have already called for a paradigm shift, and, if we want to achieve critical objectivity, historians have ready material to compare paradigms. Anderson’s argument is twofold. The first branch reveals the benefits of studying cultures in their own terms. This is useful for Western scholars studying non-Western worldviews, a category which includes historical antecedents. The second argument criticizes Western scientific rationality. Contemporary science still bears the dualist metaphysics of its origin in European modernity, and its proponents assume that, because it straddles the globe, science as currently conceived is the most successful human enterprise – and thus the standard by which to judge others. Without dismissing the benefits of modernity, the reader may decide whether a world near catastrophe is humanity’s best effort. By introducing the concept of ‘worlding’ from the humanities, Anderson’s essay opens the scientific mind to rethinking what constitutes knowledge and its applications in engineering, medicine and other domains.

A microscopic study of behavior follows Anderson’s macroscopic perspective. By examining multiple automata – chess pieces in a virtual game – David Kofman, Guillermo Campitelli and Michael Levin demonstrate how seemingly goal-oriented action can arise from multiple agents operating with limited visibility and broad autonomy. For context, I direct readers to the work of Levin and his collaborators on bioelectricity, embodied cognition and collective intelligence.

Levin, Campitelli and Kofman provide a hinge into next section, which discusses how thermodynamic entropy applies to biological systems. Again, we move from a macroscopic to microscopic perspective. Maël Montévil and Marie Chollat-Nemy offer a critical analysis of free-energy principle (FEP) and its potential role in directing cognition across life. They identify the limitations of treating organic cognition as an optimization algorithm and, thus, as a process similar to machine learning. Then José Manuel Nieto-Villar, Mariano Bizzarri and Ricardo Mansilla demonstrate how measuring the entropy production rate of a tumor can be used to judge its malignancy. Both articles treat valuable subjects, but they also serve broader purposes: they probe the limits of mathematical approaches to biology; they establish concepts from thermodynamics within biology; and they offer alternatives to current methods of mathematical modelling.

Forthcoming in early 2026, the next section turns to quantum mechanics. Here, the authors apply theoretic lenses such as affordances and Kantian Wholes to quantum theoretical constructs. They extend quantum tools to new domains, advancing the development of quantum-like frameworks for analyzing complex systems, and they argue that biological phenomena, notably evolution and the production of novelty, require different, complementary principles to those that govern physics and chemistry.

Acknowledgements

A fellowship at the Institute for Advanced Study of Nantes (*L’Institut d’études avancées de Nantes*) made this special issue possible. I would like to extend my gratitude to colleagues at the IEA-Nantes, and also to Randolph Kent, Emma de Angelis and *RUSI Journal* for their early support of my research.

Organisms

SAPIENZA
UNIVERSITÀ DI ROMA

Special Issue, “What AI Can Learn from Biology”

Vol. 8, No. 1–2 (2025)
ISSN: 2532-5876

DOI: 10.13133/2532-5876/19126

AI in This World and the Next

Greg Anderson,[a]

[a] *Department of History, College of Arts and Sciences, The Ohio State University, Columbus, United States of America*

***Corresponding author:** Greg Anderson, Email: anderson.1381@osu.edu

Abstract

As the symptoms of our self-inflicted planetary emergency become ever more alarming, hope seems to be growing that AI technologies can make our capitalist way of life more sustainable. Some even believe that machine intelligence will avert impending catastrophe more or less by itself. But the evidence of history should caution us against such heady Promethean optimism. Millennia of human experience suggest that only radical systemic change can halt our perilous trajectory. AI interventions and other such modern techno-fixes will simply not be enough.

An exciting new theoretical paradigm in the humanities and social sciences can help us grasp the full urgency of this message from history. Briefly stated, it recasts reality itself as a variable relational effect, one that humans co-produce with non-humans in the course of their everyday life practices. And just as practices have varied widely over time and space, so life has come to be experienced in a “pluriverse” of many different worlds, not in a universe of just one. An alternative pluriversal vision of history then allows us to identify striking correspondences between the sustainability of communities and their particular ways of “worlding”.

Most immediately, one can correlate the consistent sustainability of non-modern communities, past and present, with their commitment to living by a common set of metaphysical principles or “laws of being.” In stark contrast, the technoscientifc capitalist world of our own modernity, a world that current AI practices are hard-wired to perpetuate, directly violates all of these same tried-and-tested laws. The dire ecological consequences for the planet are now all too plain to see. It is vital that we learn lessons from the vast inventory of non-modern experiences and commit to re-engineering our way of worlding along more ecologically reponsible lines. Modified forms of AI can absolutely help us to realize a more livable future world in practice. But they cannot save us all by themselves.

Keywords: AI, Anthropocene, polycrisis, pluriverse, worlding, laws of being

Citation: Anderson, G 2025, “AI in This World and the Next”, *Organisms. Journal of Biological Sciences*, vol. 8, no. 1–2, pp. 7–20. DOI: 10.13133/2532-5876/19126

> We are on the brink of an irreversible climate disaster. This is a global emergency beyond any doubt. Much of the very fabric of life on Earth is imperiled. We are stepping into a critical and unpredictable new phase of the climate crisis. … We find ourselves amid an abrupt climate upheaval, a dire situation never before encountered in the annals of human existence. (Ripple *et al.* 2024, p. 1)

1. Cometh the Hour, Cometh the Techno-fix?

On October 8, 2024, an international team of experts published the latest "state of the climate report" (Ripple *et al.* 2024). It opens with the chilling passage quoted above, echoing other recent assessments (Milman 2023; Jaynes 2024). No thoughtful person can ignore the existential threats we face in this time many now call the Anthropocene. In these dire circumstances, the most urgent question we can ask about the nature and value of AI is surely: Can intelligent machines save us? As icecaps melt, sea levels rise, storms intensify, and biodiversity continues its alarming decline, can AI somehow help us resolve our planetary polycrisis?

Predictably, tech industry titans are bullish about AI's heroic potential. Kenneth Schmidt, the former Google CEO, is willing to bet that it will eventually "solve the problem" of climate change altogether, despite its own escalating environmental costs (Niemeyer & Varanasi 2024). And in wider industrial, policymaking, and academic circles, there seems to be a growing hope that AI applications can help set us on a path towards sustainability. Apparently, sophisticated imaging and mapping tools can now be used to track environmental degradation processes, like deforestation, the shrinking of glaciers, and the pollution of airs, waters, and soils. Emerging new platforms can detect carbon emissions, identify recyclable items in landfills, and increase energy grid efficiency. At the same time, drones and data management programs can help agribusiness to predict the weather, monitor soil conditions, and optimize the use of water, seeds, herbicides, and other resources (Flanagan 2024; Masterson 2024).

But what if such techno-fixes are not enough? What if genuine sustainability requires us to do more than curb the excesses of our modern way of life, maximize its efficiencies, and mitigate its more catastrophic effects? What if, after all the damage already inflicted upon them over the past few hundred years, Earth's fabrics just cannot take too much more of our modernity, however tempered in form?

The authors of the 2024 "state of the climate" report are not alone in believing that more radical change is urgently needed, not least because capitalism's core commitment to "unlimited growth" is self-evidently a "perilous illusion" (Ripple *et al.* 2024, p. 10). But what might a more ecologically responsible way of life actually look like in practice? And how might AI help us to negotiate the transition to this more sustainable order?

These are the questions I wish to explore in the rest of the paper. My ultimate aim is to broaden the horizons of current discussions around AI and the polycrisis by drawing on the ample resources of history, with some help from anthropology, critical theory, and "traditional ecological knowledge" along the way.

For the historical record offers a forceful corrective to any faith in the power of modern technologies to resolve our planetary predicament by themselves. As it reveals, there is a remarkably strong consensus among non-modern peoples, from prehistory to the present, about the basic kinds of truths that humans must abide by if they are to live with Earth, not against her. And these non-modern truths are diametrically opposed to those which anchor our whole modern technoscientific capitalist way of life.

Of course, taking this tried-and-tested wisdom of the ages seriously requires us to suspend our own modern common sense, which would tend to dismiss non-modern ways of knowing as "primitive" and "unscientific". But if we are prepared to make this effort, our whole way of thinking about planetary life in the past, present, and future will be duly transformed. We shall see why growing numbers of influential authorities now believe that humans have always lived in a "pluriverse" of many worlds, not in a universe of just one. And from this alternative pluriversal perspective, we can begin to view both AI and the polycrisis in productive new ways.

2. Alone Together in a Pointless Universe

Before we can traverse history's pluriverse, we need to reconsider the modern universalist common sense that would prevent us from getting there in the first place. The following account summarizes the one-world reality that is baked into the political, social, economic, legal, educational, and other core mechanisms of our modern way of life, AI included (Mignolo 2011; Descola 2013; Anderson 2018).

In the modern West, we are socialized to think of reality as a more-or-less boundless universal space, a cosmos without axial center or fixed limits. Simultaneously everywhere at once and nowhere in particular, it is just a vast container of multitudes of discrete individuated entities, all defined by their own innate properties and existing ultimately for themselves. Indeed, this modern universe seems to have no larger animating purpose or meaning beyond its own all-inclusive universality. It may be governed by machine-like physical "laws" which produce recurring patterns among its contents. But it is not at all clear where these impersonal laws come from, why they do what they do, or what ultimate ends they serve. Our cosmos just arbitrarily exists for itself. Devoid of animating aim or intentional design, it is just a pointless play of things and forces in otherwise empty space.

To qualify as real in this clockwork universe, things must be reducible to materialities that are observable to humans, whether they be directly visible material things, like sand grains, persons, and planets, or things that are detectable through their perceived material effects, like atoms, gravity, and wind. Our reality thus excludes unobservable things that seem to defy nature's physical laws, like gods, demons, and other "supernatural" phenomena. In the end, such things depend for their existence on the human mind, as subjective products of personal beliefs. To qualify as real, a thing must exist objectively, as a materially self-evident mind-independent entity.

So which things in this objectively knowable world are the most important? The short answer is human beings. Humans in our reality are always exceptional. Like other things, we humans are programmed to function as free-standing self-realizing entities, to stand for ourselves as individuals. But *unlike* other things, we are also born with personhood, which gives us special properties like consciousness, reason, language, agency, and rights to life, liberty, and property. In other words, we humans are the only true subjects in a universe full of objects. We are not accountable to any other-than-human persons, since no such beings truly exist. And we alone can judge what is real, since we alone can know the world objectively, viewing it as if from outside, like gods.

As a result, our reality inevitably resolves itself into two distinct orders: a higher order of "culture" that contains exclusively human things, like persons and cities, societies and economies, arts and sciences; and a lower non-human order of "nature", which is merely an "environment" of impersonal automata and mechanical processes. With our property rights and our freedom from accountability to non-humans, we humans can thus exploit the natural order however we want.

This vision of a secular material world dominated by free human individuals duly shapes our preferred modern way of life, with its democracies, its capitalist economies, and its rights-based notions of citizenship. If we humans are programmed to live ultimately for ourselves as rational, acquisitive, self-actualizing beings, it makes sense to order our lives in ways that will allow such beings to thrive and prosper. It makes sense to separate off a "sacred" sphere of irrational belief in gods from a "secular" sphere, where all the real business of life can be rationally transacted. It makes sense to use forms of government that grant all human subjects their right to self-determination. Yet it also makes sense to confine this government within its own realm of "public" power, sealing it off from the "private" realms of society and economy, where individuals can be free to act on their natural instincts to manage and enrich themselves.

We tend to take this account of a materialist, anthropocentrist, secularist, and individualist reality for granted, not least because it is hard-wired into all the structures that govern and define our whole modern way of life. And one might suppose that the objective truth of this account has been "proved" by the success of that way of life over recent centuries, with all its technological innovations, complex societal systems, and vast accumulations of aggregate wealth. But history suggests otherwise.

For it is undeniable that countless non-modern peoples across time and space have successfully staked their lives on accounts of reality that are profoundly different from our own, flourishing on their own terms for hundreds, sometimes even thousands of years. Moreover, unlike ourselves, they have consistently managed to thrive in ways that seem to have been sustainable, without imperiling the whole future of the planet in just a few hundred years.

So how is it possible for humans to live successfully by dramatically different accounts of "the real world"? Could it be that reality itself is somehow plural and variable, not singular and fixed? To answer these questions, we now turn to "material semiotics", a recent current in critical theory that can help us to re-visualize the human story in pluriversal terms.

3. Relational Being

Just as linguistic semiotics maintains that words derive their meaning from the assemblages (sentences, paragraphs, etc.) in which they are embedded, material semiotics proposes that entities derive their being from their relations with other entities. Whenever networks or "webs" of persons and things are collated by our life-sustaining practices, their human and non-human components "enact each other" into reality as "actors", as things that can "make a difference" (Law & Mol 2008, p. 58). In other words, contrary to our objectivist common sense, there is no such thing as a materially self-evident thing-in-itself. Things are effectively made of their relations with the other things that make their existence possible in the first place.

To illustrate, a well-known case study shows how a sheep could be enacted as multiple different realities during a 2001 epidemic of foot and mouth disease in Cumbria, UK (Law & Mol 2008). One such sheep reality was the "veterinary sheep", a living organism that was an object of clinical examination as a site of possible disease symptoms. Another was the "epidemiological sheep", a statistical calculation based on models of infection probabilities. A third was the "economic sheep", a market-based accounting of the epidemic's impact on meat exports and on compensation claims made by farmers to the EU. And the fourth was the "farming sheep", a named member of a particular flock that stirred feelings of care and affection in its owners.

Common sense may tell us that this is just four different ways of looking at one single sheep reality. But as the authors of the case study stress, these are four different realities that are being enacted through four different webs of practice. The four sheep are ontologically distinct from one another and not always mutually reinforcing.

> You cannot learn what a sheep is by staring at a picture. It helps more to unravel the practices in which sheep figure, in which they are enacted in one way or another. If we do this then we do not discover a sheep that is unitary and coherent. Instead, we find a "sheep multiple". [T]he stories of different versions of the Cumbria sheep in 2001 both exclude and include each other. The farming sheep was invaluable, outside value, whereas the economic sheep had a price on its head. The farm flock deserved protection, whereas the economic sheep was more valuable dead than alive. And the epidemiological and the veterinary sheep clashed with and depended on one another (Law & Mol 2008, pp. 65-66)

When more generally applied, this rigorously relational way of accounting for the contents of experience can thus liberate us from the black-and-white rigidity of modernity's objective world. It enables us to tell stories about reality's ongoing constitution that are dynamic and fine-grained, without reducing the complex messiness of lived experience to, say, an abstract microphysics of invisible particles. Instead, by focusing on the patterned world-making interplay of persons and things, it allows us to convey a richer, more vibrant sense of the entangled abundance of being. It helps us to see reality as something fluid and elastic, as something continually in formation, not something predetermined or fixed.

In the process, material semiotics effectively rules out the possibility of a mind-independent objectively knowable world. Instead, it gives us a precise and relatively concrete way to understand how human knowledge is unavoidably implicated in the process of reality formation. If we are all necessarily embedded in a world of enacted actors, as both participants therein and products thereof, our ways of knowing that world will always be historically situated. What we know and how we know it will inescapably be conditioned by all of those beings and things which enact us as knowledgeable actors in the first place. And reality will then be the complex ongoing *effect* which is generated whenever that knowledge and the world appear to be in alignment. Which is to say, reality is the enacted effect of a mind-independent world, not its literal actuality.

If so, there can be no single absolutely or universally "true" or "right" way of knowing what's really there, because everything is potential multiplicity and what counts as knowledge will always be historically mediated. What matters, then, is not that our knowledge conforms to some timeless abstract truth standard, objective or otherwise. What matters is that the world which our knowledge predisposes us to enact is actually realizable and hopefully sustainable in practice, whether we are, say, ancient Egyptians, Indigenous Amazonians, or modern Europeans.

4. Enacting Worlds

To describe the process of realizing the effect of a self-evident world, some now use the term "worlding". Here is way to think about it.

Every human community stakes its life on certain truths about the essential contents of experience, on shared certainties about, say, the nature of personhood and humanity, about how to relate to non-human others, about the fabrics of the lived environment and how they came to be there, and about the sources, means, and ends of life itself. As these truths become tried and tested in practice, they harden into common sense laws of being, a kind of metaphysical "model" of the world to live by. This model duly becomes embedded in the minds and bodies of community members, in all their life-sustaining norms and practices, and in their built environment, shaping their relations with one another and with all the non-humans on whom their existence depends, from animals and plants to soils and weather systems. So long as those non-humans continue to cooperate in more or less stable, predictable ways, then the community will be able to reproduce itself successfully across the generations. And the model will thus come to be continually enacted in everyday experience by humans and non-humans. In short, a worlding process produces the ongoing effect of a materially self-evident reality, a world that already seems to be there all by itself.

Hence, when the planet's non-human constituents collaborate with radically different ways of worlding, ontologically different realities are produced, as the following examples illustrate.

In classical Athens (480–320 BC), the supreme force that governed annual yields of grain and other crops was an immortal female person. The Athenians called her Demeter. Though Demeter herself was not literally visible "in the flesh" *per se*, no-one doubted her real existence in immediate experience. From childhood on, all Athenians were socialized to trust in her miraculous powers. The built environment was full of references to her significance, in poems, paintings, statues, shrines, and, above all, her sanctuary home at Eleusis. And the rhythms of each year were punctuated by gift offerings to her at great festivals like the Thesmophoria and the Eleusinian Mysteries, whereby the Athenians hoped to induce her to act favorably towards them. In return, more often than not, the goddess caused crops to grow and humans to thrive, thereby continually confirming the self-evident truth of her management of life itself.

In the modern United States, the supreme force that governs the material well-being of all humans is an impersonal machine-like system. The Americans call it "the economy". Though the economy itself is not literally visible "in the flesh" *per se*, no-one doubts its real existence in immediate experience. From childhood on, all Americans are socialized to trust in its miraculous powers. The built environment is full of references to its significance, in books, journals, news media, factories, banks, and, above all, its special home in Wall Street. The rhythms of each year are punctuated by adjustments to taxes, budgets, and interest rates, whereby the Americans hope to induce the economy to act favorably towards them. And in return, more often than not, it causes fortunes to grow and at least some humans to thrive, thereby continually confirming the self-evident truth of its management of life itself.

In these examples, Demeter and the economy are not pure constructs of the imagination. Nor are they real in any universal or absolute sense as materially self-evident things-in-themselves. A machine-like economy would be unthinkable in classical Athens, just as a superhuman goddess would be unreal in modern America. But through certain specific worlding practices, both can be enacted into existence as actors, as entities that make real differences to life itself. And once we can see reality in these relational terms as an ongoing enacted effect, history's extraordinary pluriverse of worlds can start to materialize before our eyes.

One might add a few further remarks to help us visualize this world of many worlds with a little more clarity and precision.

First and most general, one should not think of the worlds of a pluriverse as fixed, closed systems, all hermetically sealed off from one another, like a multitude of planets scattered across a firmament. As enacted effects of inherently variable life-sustaining practices, worlds themselves are inherently mutable. They can evolve, expand, contract, interact, and influence one another, The boundaries between them will always be potentially porous and plastic in principle.

Second, while worlds will almost always be anchored in particular life-nurturing terrains or habitats, the spaces they occupy need not be physically continuous or mutually exclusive. One thinks, for example, of the one thousand or so *polis* microcosms of the classical Greek cosmos, which were dispersed across vast distances between Spain and the Black Sea. At the same time, a given portion of, say, the Amazon rainforest could simultaneously be two different things in two different worlds. It could be enacted both as a parent-like home by local Indigenous communities and as an inert bundle of economic resources by capitalist corporations.

Third, the worlds of a pluriverse need not be internally monolithic. While the overall metaphysical temper of a world will be established by the laws of being that are baked into the routine practices of the majority or dominant group, there may still be room within for alternative ways of worlding by minority or subordinate constituencies, thereby complicating the fabrics of the whole.

For instance, both the Roman and Chinese empires at certain times accepted that some subjects would maintain relations with alien gods, divinities whose presence in the worlds in question was not officially recognized. But such internal variations are perhaps most readily visible in the world of modernity itself. Yes, lives may now be almost universally staked on political, economic, legal, educational, and other mechanisms that enact a modern materialist, anthropocentrist, secularist, and individualist cosmos into being. But during the Cold War era, for example, one could still identify ontological differences between "capitalist" and "communist" versions of modernity, not least in their respective enactments of the "free market economy" and the "Communist Party" as the supreme world-making agencies. And even today, to a point, it seems reasonable to speak of different national microcosmic modernities across the globe, especially where vestiges of non-modern worlding practices remain. But while these counter-worldings may give the fabrics of everyday being a certain distinctively local or regional coloring, they do not fundamentally change those same essential fabrics.

Fourth, worlds will change and evolve as the laws of being embedded in worlding practices change and evolve, whether the causes are internal or external. Such changes were triggered, for example, by the processes we call the "Christianization" of the Roman empire and the British "colonization" of South Asia. In both of these cases, a counter-worlding project ultimately prevailed because it was imposed from above and backed by force, fundamentally altering what would count as reality and the very meaning of life itself. And external pressure for such change continues to this day to disrupt what survives of Indigenous ancestral worlds, almost all of which have been complicated to some degree by modern ways of worlding, inevitably rendering them somewhat "hybrid" in nature as a result (Halbmayer 2018).

5. The Wider Stakes

Radical as it may seem, this alternative many-worlds vision of reality is no longer an eccentric or fringe proposition. Though attempts to theorize the worlding process may vary slightly in their particulars, a general commitment to pluriversal thinking has been embraced by growing numbers of authorities in a range of different fields, including anthropology, history, international relations (IR), decolonial theory, and science and technology studies (STS). There are several mutually reinforcing reasons for making this commitment.

As prominent STS authors have shown, one can make a robust case for a pluriversal alternative on purely theoretical grounds, using material semiotics and/or other related critical currents (Law 2015). Then again, as specialists in anthropology, history, and IR have demonstrated, a case can also be made on the grounds of analytical utility, since one can only make meaningful sense of history's many ways of being human if one understands each one on its own ontological terms, in its own local world of experience (Holbraad & Pedersen 2017; Anderson 2018; Schaarsberg 2023). Nor should we overlook the ethical case for pluriversal thinking, which would insist that all peoples across time and space, especially today's Indigenous communities, should have the power to determine the ultimate truths of their own existence (Escobar 2017; Anderson 2018). But perhaps the most fundamental reasons for embracing a many-worlds vision of reality are not philosophical or academic at all. They are ecological, even existential. After all, the potential stakes could hardly be higher.

To begin with, a pluriversal perspective allows us to see that the human causes of our current polycrisis are not just to be found in particular modern practices, like those associated with carbon emissions, industrial pollution, and the loss of biodiversity. They are ultimately to be found in particular laws of being which have rendered those same practices normal, acceptable, even natural over time. In other words, these causes are endemic to an entire way of worlding, to a historically unprecedented way of being human that people of European descent have exported around the globe over the past few hundred years, often destroying other more sustainable worlds in the process. Among the many thousands of different worlds in history's wondrous pluriverse, only this modern kind has metaphysically prioritized the material over the ideational, the human over the non-human and the superhuman, ultimate knowability over ultimate mystery, and the life of the individual over that of the social body. The net results of this way of worlding are now all too clear to see.

At the same time, a many-worlds vision of the human story can also exponentially enrich our quest for more sustainable alternatives, inviting us to learn from a vast horizon of worlds which have been far more ecologically balanced than our own.

6. Five Historical Laws of Being

Non-Indigenous biologists and ecologists have long been demonstrating the practical utility of "traditional ecological knowledge" (TEK) through case studies in various parts of the globe (Johnson 1992; Berkes *et al.* 2000). But latterly, this subject area has been reclaimed by Indigenous authorities, who are far better placed to explain why bodies of TEK are consistently effective in practice (Cajete 2000; Nelson & Shilling 2018). To this large inventory of evidence, one can add all the life-sustaining wisdom that has been recovered by historians and others who study peoples of the past. When we then survey all this non-modern know-how, some significant patterns emerge. The following five common laws of being help to explain the consistent sustainability of non-modern worlds.

a. Being is belonging

All being is local. Every known non-modern world is a concrete somewhere not a universal everywhere. It is always defined and conditioned by a specific habitat, a nurturing parent-like cradle of life to which it is congenitally attached. And across history's pluriverse, these home environments have taken many different forms.

For example, forests have been the world-defining providers of all life's needs for peoples like the Mbuti of the Democratic Republic of Congo, the Kajang of Indonesia, the Nayaka of southern India, the Yanomami, and numerous other Indigenous Amazonians (Kopenawa & Albert 2013). People of Quechua descent in the Peruvian Andes may relate to mountains like Ausangate as apus, the fatherly counterparts to *pachamamas*, the life-giving mothers of the earth (Carreño 2016). Maori *iwi* likewise relate to great rivers like the Waikato and Whanganui as parental sources of vitality (Salmond *et al.* 2019). And for the boat-dwelling Badjao people, a similarly nurturing role is performed by the seas around the Philippines and Indonesia (Macalandag 2023).

But of all the diverse habitats with which humans have maintained kin-like relations over the centuries, land itself is of course by far the most common. In some worlds, like those of the classical Athenians, the Hopi, Zuni, and other Native peoples of the United States, the first humans literally emerged from a womb-like Mother Earth (Anderson 2018; Homburg *et al.* 2023). In other creation stories, the original humans are partly or wholly made from earthy materials, as we see in the Book of Genesis, the Qu'ran, the Mesopotamian *Atra-Hasis* epic, and the ancestral traditions of the Dayak of Borneo, the Vietnamese, the Malagasy, and the Inka.

What is common to all these instances is a profoundly un-modern sense of consubstuntiality or continuity of being between humans and their habitats. Whether they know themselves as offspring of an earth mother or as creatures made directly from home terrains, most if not all non-modern peoples have experienced a sense of environmentally embedded belonging that rules out any possible nature/culture divide.

Also unthinkable would be the idea of a universal world without center or limits. Non-modern worlds almost invariably gravitate around a fixed focal point, an *axis mundi* from which vital energies radiate out across the cosmos, unifying the whole. These axial points may be "trees of life", like the Norse Yggdrasil and the Mayan Yaxche. They may be "holy mountains", points of contact between terrestrial and celestial realms, like the Daoist Kunlun and the Black Hills of the Lakota. They can be centripetal sites of ritual activity, like the Javan Borobudur and the Hebrew temple in Jerusalem.

And in imperial worlds, cities can perform this role, like Rome and Constantinople, Babylon and Mecca, Nanjing and Beijing, Cusco and Tenochtitlán.

Furthermore, non-modern worlds are always finite in practice, with habitats defining both their physical and metaphysical limits. This means there is a constant sense of insecurity among non-modern humans, because life's sources are inevitably exhaustible. As a result, some of the most inviolable rules which non-modern peoples live by are those which limit the use of vital resources, preserving them for all generations to come. Under such conditions, the idea of staking one's well-being on a vision of "unlimited growth" would be wholly self-defeating.

b. A world is a symbiotic ecology

Non-modern worlds are never mere containers of disaggregated subjects and objects. On the contrary, they tend to be self-reproducing symbiotic ecologies. All their component parts, both human and non-human, are thus effects of their mutually dependent relations with others.

For example, the ancient Athenian *polis* was a cosmic ecology, where life was sustained by ongoing collaborations between the Athenian people, their divine motherland of Attica, and the two hundred gods who furnished all their other conditions of existence, from sunshine and rainfall to human health and battle outcomes (Anderson 2018). In the medieval European Great Chain of Being, all the contents of Creation, from stones and waters to plants, humans, and angels, were expressly designed by God to perform assigned roles in the world's perpetuation (Lovejoy 1976). In the cosmos of Ming China, the emperor, as "Son of Heaven", had a divine mandate to align all things in the earthly realm with the timeless "Way" of the celestial realm (Jiang 2011). Elsewhere, all components of the ancestral Andean world of Abya Yala, from the smallest pebbles to *pachamamas*, are active beings who contribute to the healthy balanced life of the whole (Amawtay Wasi 2004). And in the microcosmic worlds of Maori *iwi*, humans and non-humans are kindred descendants of the same *whakapapa*, an all-inclusive multi-species genealogy (Harmsworth & Awatere 2013).

Hence, in these and other non-modern realities, the human person itself is always in some sense a relational being. There is no such thing as a modern-style self-actualizing individual.

In some worlds, interdependent community members can routinely act with the mind, will, and interest of a single indivisible person, like the *demos* of the Athenians, the Roman *populus*, or a medieval "body politic". A unitary corporate person of this kind always precedes and outlives all the living breathing humans who embody it at any given time. Likewise, the Ming empire's vast body of government officials served as extensions of the mind and body of the emperor himself when furthering his work of mediation between heavenly and earthly realms (Jiang 2011). Elsewhere, the divine king of the precolonial Hawai'ians could "encompass the people in his own person, as a projection of of his own being" (Sahlins 1985, pp. 207, 214).

More common are worlds where each human is enacted as a "dividual" person, a composite of life-defining elements that derive from relations with others. Among the Dogon of Mali, each person is composed of three elements from different sources: a physical body (*goju*) from the father; a character (*hakile*) from the mother or father; and an inner vitality (*kikine*) from the creator god Ama (van Beek 1992). In a traditional Hindu world, a person is a more permeable and fluid being, an ongoing coalescence of substances that are exchanged in one's relations with others, like blood, cooked food, money, words, and knowledge (Marriott 1976). And for the Hagen of Papua New Guinea, every person is a "social microcosm", a "plural and composite site of the relations that produced them" (Strathern 1988, p. 13).

c. Humans are not alone

Humans are never alone in non-modern realities. They always share life's experiences and responsibilities with communities of other-than-human persons.

In many cases, like those of ancient Greece, Rome, Egypt, Persia, China, and Hindu South Asia, the most important of these non-human persons are gods and other immortal beings. Though usually invisible, these numinous agencies are actively present in immediate experience. They do not inhabit some otherworldly elsewhere, leaving Creation to run itself. They continually manage the infrastructure of the cosmos, being immanent in its celestial bodies, soils, rivers, and other fabrics. Their personal wills thus control all of life's conditions, sources, processes, and outcomes. And humans continually seek their favor, socializing with them in their sanctuaries and other special haunts.

In numerous other non-modern worlds, a more diverse array of other-than-humans share a human-like consciousness, agency, and subjectivity. In the ancestral world of the Sámi in Fenno-Scandia, things like land, forests, lakes, rivers, fish, and reindeer all have their own personalities (Helander-Renvall 2010). In the cosmos of the Chewong of Malaysia, "our people" (*bi he*) includes all things, from spirits to animals and plants, that possess *ruwai* or "reflexive consciousness" (Descola 2013, pp. 26-27). For the precolonial Lakota, the world teemed with "all my relatives" (*mitakuye oyasin*), including animals who lived in their own human-like "nations" (*oyate*), "lodges" (*tiyospaye*), and households (Posthumous 2017). And in the worlds of Amazonians like the Makuna, each animal species enacts the human role in its own microcosmic reality, complete with its own shamans, rituals, houses, fermented drinks, and so on (Viveiros de Castro 1999).

d. Life demands accountability to others

Life in non-modern worlds therefore depends on collaborations with a host of other-than-human persons. It thus brings with it duties of care, respect, gratitude, and accountability towards those others, if the symbiotic ecology is to remain in equilibrium.

In worlds governed by pantheons of divinities, the human obligation to show care, respect and accountability to those others may be discharged through, say, prayers, sacrifices, votives, and invitations to gods to participate in rituals. Of course, conventional academic wisdom tends to see all such activities as mere exercises in "religion", as expressions of an ultimately irrational, subjective belief in the existence of unreal "supernatural" beings. But in worlds where gods control all the material conditions of existence, such practices are not just entirely rational. They are life-sustaining ecological mechanisms. Only by maintaining positive relations with the managers of the cosmos through ritual actions can communities hope to flourish.

In worlds where personhood is more widely dispersed among the contents of Creation, the practice of accountability to others assumes an even wider range of different forms. For example, when engaging in lake fishing, Sámi should abide by an ethic of *jávrediksun*, a sense of responsibility for the long-term well-being of both the lake and its fish (Østmo & Law 2018). To ensure that caribou willingly give themselves to sustain human lives, the Innu of Labrador commit to sharing their meat appropriately, treating their other body parts with respect, and maintaining good relations with Kanipinikassikueu, the caribou spirit master (Blaser 2016). Similarly, shamans of the Amazonian Makuna must engage in ongoing negotiations with the spirit masters of other species over the animals and fish they hunt, making offerings to ensure that lost lives are replaced (Arhem 1996).

e. Experience is ultimately mysterious

If all non-modern peoples thus accept humanity's relatively humble place in the cosmic order, they also accept limits on human abilities to know that order. They all must coexist with other-than-human persons who know things that humans could never know. And they all must live among invisible beings and forces that are, by definition, beyond human understanding.

To be sure, the mysterious wills of the cosmos may be divined by humans with extraordinary aptitudes or special ancestries, like Egyptian temple astrologers, the Pythia at Delphi, Amazonian shamans, and the *babalawos* of Afro-Cuban Ifá. And many peoples have learned things from visible other-than-humans, like trees, plants, animals, birds, and waters. For them, as Lakota Chief Luther Standing Bear once said, Creation is an inexhaustible "library" of knowledge (Standing Bear 1976, p. 194). But in all these cases, there are also things that are just not for humans to know. In all these cases, the idea of an objectively knowable universe would be arrogantly presumptuous if not utterly delusional.

In short, the evidence of a pluriversal history offers an implicit critique of our whole modern way of worlding. Modernity's materialist, anthropocentrist, secularist, and individualist laws of being have not just departed from all historical norms. They consistently violate the principles that have allowed humans to thrive sustainably across the millennia. The basic lessons that non-modern peoples teach us are thus clear enough. Instead of forcing planetary life to align with human priorities, we need to force our priorities to align with planetary life. We need to recommit to ways of worlding that are more locally grounded, more symbiotically relational, and more humbly sensitive to all the other-than-human conditions of our existence.

7. Worlding Against the Modern Grain

Daunting a challenge as this may seem, it is important to know that many communities around the globe are already pursuing ways of life along these more ecologically balanced and sensitive lines. Some of these counter-worldings are happening in remote locations, like jungles and tundras, continuing ancestral practices of yore . But others are newer projects, evolving even in the heart of major cities in the Global North.

For a start, there are still many surviving "territories of life", where local communities are actively working to maintain time-tested non-modern ways of worlding, sometimes with financial and other support from organizations like the UN Equator Initiative, the ICCA Consortium, and La Via Campesina (Borrini-Feyerabend 2024). These communities range from "foragers", like the Wampís Nation of Amazonian Peru, to "mobile pastoralists", like the Sarikeçili Yörüks of Turkey, to "shifting cultivators", like the Kavet of Cambodia. And they include western European groups, like the female shellfishers-on-foot (*mariscadoras*) on Spain's Galician coast and the guardians of the Regole d'Ampezzo in Italy, who manage their alpine ecological enclave according to original medieval prescriptions.

Nor can we ignore the ongoing resistance to settler colonialism by many Indigenous communities, who have been struggling to reclaim their ancestral lands and their right to determine for themselves what counts as a world. Such decolonial struggles have become increasingly prevalent since the later 1960s, seeking liberation from a modern way of worlding that casts Indigenous peoples as a perennial "problem" for capitalist "development" (Clifford 2013). In recent decades, countless groups and communities have pursued decolonial causes: from the Mapuche in Chile and Zapatistas in Mexico to the Innu and Inuit in northern Canada; from the Sámi of Fenno-Scandia to the Yakuts of Siberia and Itelmen of Kamchatka; and from the Noongar and other First Nations in Australia to dozens of Maori *iwi* in Aotearoa New Zealand (Bauer 2021; Dewar 2009; Sulyandziga & Berezhkov 2023; De Villiers 2020).

To all this, one should add the proliferation of new efforts to pursue alternative ecological pathways around the globe, even in Europe and the United States. For example, the "social solidarity economy" in Catalunya, Spain, now involves some 140,000 workers in over 7,000 organizations, including co-ops, mutual aid societies, and exchange networks (Lees 2022). Among many rurally-oriented "degrowth" initiatives in Europe is Cargonomia in Hungary, which uses a fleet of cargo bicycles to deliver local organic produce directly to customers (Lorenzen & Moore 2022, p. 48). In Mississippi, the African-American-led Cooperation Jackson seeks "sustainable community development" through various worker-owned ventures, treating land as an active "partner", not as inert "property" (Akuno & Meyer 2023). And such projects are now supported by a host of national and international organizations, like the Black Land and Liberation Initiative in the United States, the Chantier de l'Économie Sociale in Canada, and the International Network for the Promotion of Social Solidarity Economy.

More generally, there seems to be an increasing willingness in the wider environment to question some of the common sense that underpins our modern way of worlding. One sees a growing interest in locally embedded, "bioregional" alternatives to globalizing capitalism (Bove 2021). The "rights of nature" cause, which seeks to establish legal personhood for a range of different non-humans, has become ever more mainstream across the planet since the 1970s (Stone 1972; Surma 2021; Bosselmann & Williams 2025). Meanwhile, scientists now commonly subvert the nature/culture divide by attributing forms of cognition, intelligence, subjectivity, and sociality to all manner of other-than-humans, including animals, micro-organisms, fungi, plants, trees, and rivers (Bouteau *et al.* 2021; Simard 2021; Calvo 2023). And it is no less commonplace to recognize that collaboration, mutualism, and symbiosis are essential to vitality at all scales, from the cellular to the planetary (Margulis 1998; Weiss & Buchanan 2009; Bronstein 2015). Why should human vitality be any different?

Needless to say, these various forms of counter-worlding are not yet sufficiently prevalent or influential to remake the fabrics of modern being from within. By themselves, they cannot secure a transition towards the more relationally grounded, more ecologically responsible, more pluriversal world of the future that our planetary crisis seems to be demanding. Nonetheless, these diverse oppositional causes and projects do at least help us to visualize such a shift, giving us a more concrete sense of what more sustainable ways of worlding might actually involve in practice. Formidable as the obstacles to radical change may still be, movement in this direction is already happening, if we are only willing to see it.

8. AI in This World and the Next

What then might this alternative pluriversal perspective reveal about the nature of AI and its capacity to support more relational, more sustainable ways of worlding? Here are three closing thoughts.

First, while AI as we know it surely can help to mitigate some of the more overt symptoms of the polycrisis, it would be folly to bet on the faint hope that it might somehow reverse our catastrophic trajectory altogether. Right now, one could argue, it is more part of the problem than the solution. As even the most ardent supporters of AI acknowledge, its potential ecological benefits are already compromised by its troubling environmental costs (Ren & Wierman 2024; Winston 2024). But more alarming from a pluriversal perspective are the environmental *consequences* of the practices that its routine operations make possible.

By now, AI is thoroughly enmeshed as an enacted actor in myriad webs of practice. It is already making differences in almost every field of modern endeavor, from commerce and industry to communication and education. The problem is that most of these differences are reenergizing a manifestly unsustainable way of life, thereby perpetuating the delusional dream of unlimited growth. Our personal computers are bombarded with algorithm-driven advertisements that create yet more demand for all manner of goods, regardless of the planetary costs required to produce those goods and ship them to consumers. AI now commonly helps the fossil fuel and other extractivist industries to refashion ever more of Earth's fabrics into profitable commodities. Meanwhile, plagues of online bots are corrupting elections with misinformation, almost always to favour forces that are hostile to environmental controls. So even as certain AI applications may be inching us towards a more sustainable future, the ever growing complicity of other applications with capitalist "business as usual" is taking us yet further away from that goal.

Second, when we recontextualize AI in a many-worlds scheme of history, we become more acutely aware of its epistemic limitations. Today, the prospect of an all-knowing artifical general intelligence or superintelligence stirs both excitement and alarm (Kurzweil 2004; 2024; Bostrom 2014). But when viewed through a pluriversal lens, this prospect all but evaporates. Remarkable as the powers of AI may already be, it has so far internalized the knowledge of just one kind of world. In history's many other worlds, we not only see thousands of other tried-and-tested ways of knowing the grains of experience. We find profoundly different ways of determining what counts as knowledge in the first place.

As we have seen, bodies of non-modern wisdom are not just accumulations of data about, say, ancestral traditions, ritual procedures, or harvesting techniques. They are fundamentally relational ways of knowing experience. They include commitments to show care, respect, and accountability to the other existents on whom one's life depends. They include a sense of being a component part of things larger than oneself, an innate feeling of belonging to a particular habitat and its symbiotically entangled communions of humans and other-than-humans. And they fundamentally include an experience of insecurity and ontological humility, of being continually subject to higher powers, to forces and exigencies that humans can scarcely comprehend, never mind control.

None of these are things that can readily be measured, quantified, or simulated through discrete data bytes, algorithms, or computer codes. Could a machine ever truly know what it is to live in a more fluid, more open-ended world, where things are made of relations, where being is always becoming? Could a machine ever truly *care*? Many today speculate about the possibility of a "sentient" AI (Long *et al.* 2024). But the kind of machine sentience they envisage is always a facsimile of a modern human subjectivity, mechanically reproducing what it is like to be a free-standing self-actualizing individual. It is thus very hard to imagine that AI will ever acquire the oracular powers that might guide us toward the other ways of worlding that we so urgently need. Unlike the actual oracles in many non-modern worlds, a machine will never possess the kind of transcendental relational wisdom that can see what's best for the cosmos as a whole.

Third, AI applications could nonetheless play important auxiliary roles in a transition to a more sustainable future. It is not hard to see how they might be productively woven into networks of practice that are already driving counter-worlding processes, serving the needs of, say, Catalunya's solidarity economy, Hungary's Cargonomia, or Cooperation Jackson. Like their capitalist counterparts, such alternative bioregional projects would clearly benefit from advanced technological assistance with things like weather prediction, efficiency maximization, waste recycling, and the distribution of vital resources. And it is not impossible to imagine how species of machine intelligence could be used by the ever-growing host of organizations that support ancestral ways of worlding across the globe, helping them to keep track of community fortunes and dispense aid in the forms required.

In sum, AI may not be the heroic change agent that some wish for. It cannot save us all by itself. But if repurposed to serve the greater ecological good, it could still make significant differences, helping us forge our necessary passage from this world to the next.

References

Akuno, K & Meyer, M (eds.) 2023. *Jackson Rising Redux: Lessons on Building the Future in the Present*. Oakland, CA: PM Press.

Amawtay Wasi Cross-Cultural University 2004. *Learning Wisdom and the Good Way to Live*. Quito, Ecuador: UNESCO.

Anderson, G 2018. *The Realness of Things Past: Ancient Greece and Ontological History*. New York: Oxford University Press.

Arhem, K 1996. "The cosmic food web: Human-nature relatedness in the Northwest Amazon", in Descola, P & Palsson G (eds.) *Nature and Society: Anthropological Perspectives*. London: Routledge,pp. 185–204.

Bauer, K 2021 *Negotiating Autonomy: Mapuche Territorial Demands and Chilean Land Policy*. Pittsburgh: University of Pittsburgh Press.

Berkes, F *et al.* 2000. "Rediscovery of traditional ecological knowledge as adaptive management", *Ecological Applications* vol. 10, no. 5, pp. 1251–1262.

Blaser, M 2016 "Is another cosmopolitics possible?" *Cultural Anthropology*, vol. 31, no. 4, pp. 545–570.

Borrini-Feyerabend, G, with Jaeger, T 2024 *Territories of Life. Exploring the Vitality of Governance for Conserved and Protected Areas*. Genolier, Switzerland: ICCA Consortium. https://volume.territoriesoflife.org/

Bosselmann, K & Williams, T 2025. "The river as a legal person: The case of the Whanganui River in New Zealand". In: *www.boell.de*, Berlin. https://www.boell.de/en/2025/01/29/river-legal-person-case-whanganui-river-new-zealand.

Bostrom, N 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Bouteau, F *et al.* "Our sisters the plants? Notes from phylogenetics and botany on plant kinship blindness", *Plant Signaling and Behavior*, vol. 16, no. 12. https://doi.org/10.1080/15592324.2021.2004769

Bove, T 2021. "Bioregionalism: A model for self-sufficient and democratic economy". In: *earth.org*, Hong Kong. https://earth.org/bioregionalism/

Bronstein, JL (ed.) 2015 *Mutualism*. Oxford: Oxford University Press.

Cajete, G 2000 *Native Science: Natural Laws of Interdependence*. Santa Fe, NM: Clear Light Publishers.

Calvo, P 2023. *Planta Sapiens: The New Science of Plant Intelligence*. New York: W. W. Norton.

Carreño, GS 2016. "Places are kin: Food, cohabitation, and sociality in the southern Peruvian Andes", *Anthropological Quarterly*, vol. 89, no. 3, pp. 813–840.

Clifford, J 2013 *Returns: Becoming Indigenous in the Twenty-First Century*. Cambridge, MA: Harvard University Press.

Descola, P 2013 *Beyond Nature and Culture*. Chicago: University of Chicago Press.

De Villiers, B 2020 "Privatised autonomy for the Noongar People of Australia: A sui generis model for indigenous non-territorial self-government", *Verfassungs und Recht Übersee*, vol. 53, pp. 171–189.

Dewar, B 2009 "Nunavut and the Nunavut land claims agreement—An unresolved relationship". In: *policyoptions.irpp.org*, Montréal. https://policyoptions.irpp.org/magazines/canadas-water-challenges/nunavut-and-the-nunavut-land-claims-agreement-an-unresolved-relationship/

Escobar, A 2017 *Designs for the Pluriverse: Radical Interdependence, Autonomy, and the Making of Worlds*. Durham, NC: Duke University Press.

Flanagan, M 2024 "AI and environmental challenges". In: *environment.upenn.edu*, Philadelphia. https://environment.upenn.edu/events-insights/news/ai-and-environmental-challenges

Halbmayer, E (ed.) 2018 *Indigenous Modernities in South America*. Canon Pyon, UK: Dean Kingston Publishing.

Harmsworth, GR & Awatere, S 2013. "Indigenous Maori knowledge and perspectives of ecosystems", in Dymond, JR (ed.) *Ecosystem Services in New Zealand: Conditions and Trends*. Lincoln, NZ: Manaaki Whenua Press, pp. 274–286.

Helander-Renvall, E 2010 "Animism, personhood, and the nature of reality: Sami perspectives", *Polar Record*, vol. 46, pp. 44–56.

Holbraad, M & Pedersen, M A 2017. *The Ontological Turn: An Anthropological Exposition*. Cambridge, UK: Cambridge University Press.

Homburg, J *et al.* 2023 "From native tradition to modern-day America: Native origin legends that involve soil and earth", in Patzel, N *et al.* (eds.) *Cultural Understanding of Soils: The Importance of Cultural Diversity and of the Inner World*. Cham, Switzerland: Springer, pp. 181–207.

Jaynes, CH 2024 "Planet will warm as much as 3.10 C under current policies: UN Report". In: ecowatch.com, Cleveland. https://www.ecowatch.com/unep-emissions-gap-report-2024-climate-crisis.html

Jiang, Y 2011 *The Mandate of Heaven and the Great Ming Code*. Seattle: University of Washington Press.

Johnson, M (ed.) 1992 *Lore: Capturing Traditional Ecological Knowledge*. Ottawa: International Development Research Council.

Kopenawa, D & Albert B 2013 *The Falling Sky: Words of a Yanomami Shaman*. Cambridge, MA: Harvard University Press.

Kurzweil, R 2004 *The Singularity is Near: When Humans Transcend Biology*. New York: Viking.

Kurzweil, R 2024 *The Singularity is Nearer: When Humans Merge with AI*. New York: Viking.

Law, J 2015 "What's wrong with a one-world world?", *Distinktion: Scandinavian Journal of Social Theory*, vol. 16, no. 1: pp. 126–139.

Law, J & Mol, A 2008 "The actor enacted: Cumbrian sheep in 2001", in Knappett, C & Malafouris, L (eds.), *Material Agency: Towards a Non-Anthropocentric Approach*. New York: Springer, pp. 57–77.

Lees, E 2022 "The Catalan social solidarity economy". In: *ppesydney.net*, Sydney. https://www.ppesydney.net/the-catalan-social-and-solidarity-economy/

Long, R *et al.* 2024 "Taking AI welfare seriously", *arXiv* 2411.00986v1. https://arxiv.org/html/2411.00986v1

Lorenzen, H & Moore, O (eds.) 2022 Rural Europe takes action: No more business as usual. Brussels: Forum Synergies and ARC2020. https://www.arc2020.eu/wp-content/uploads/2022/07/RETA-final-5MB.pdf

Lovejoy, A 1976 *The Great Chain of Being: A Study in the History of an Idea*. Cambridge, MA: Harvard University Press.

Macalandag, R 2023 "The Badjao and the sea: Indigenous entanglement with coastal resource management—The case of 'settled nomads' in the Philippines", in D'Arcy P & Kuan DDD (eds.) *Islands of Hope: Indigenous Resource Management in a Changing Pacific*. Canberra: ANU Press, pp. 109–122.

Margulis, L 1998. *Symbiotic Planet: A New Look at Evolution*. New York: Basic Books.

Marriott, M 1976. "Hindu transactions: Diversity without dualism", in Kapferer B (ed.) *Transaction and Meaning: Directions in the Anthropology of Exchange and Symbolic Behavior*. Philadelphia: Institute for the Study of Human Issues, pp. 109–142.

Masterson, V 2024 "9 ways AI is helping tackle climate change". In: *weforum.org*, Cologny. https://www.weforum.org/agenda/2024/02/ai-combat-climate-change/

Mignolo, W 2011 *The Darker Side of Western Modernity: Global Futures, Decolonial Options*. Durham, NC: Duke University Press.

Milman, O 2023 "We are damned fools: Scientist who sounded climate alarm in 80s warns of worse to come". In: *theguardian.com*, London. https://www.theguardian.com/environment/2023/jul/19/climate-crisis-james-hansen-scientist-warning

Nelson, MK & Shilling, D (eds.) 2018 *Traditional Ecological Knowledge: Learning from Indigenous Practices for Environmental Sustainability*. Cambridge, UK: Cambridge University Press.

Niemeyer, K & Varanasi, L 2024 "Former Google CEO says we should go all in on building AI data centers because 'we are never going to meet our climate goals anyway". In: *businessinsider.com*, New York. https://www.businessinsider.com/eric-schmidt-google-ai-data-centers-energy-climate-goals-2024-10

Østmo, L & Law, J 2018 "Mis/translation, colonialism, and environmental conflict", *Environmental Humanities*, vol. 10, no. 2, pp. 349–369.

Posthumous, D 2017 "All my relatives: Exploring nineteenth-century Lakota ontology and belief", *Ethnohistory*, vol. 64, no. 3, pp. 379–400.

Ren, S & Wierman, A 2024 "The uneven distribution of AI's environmental impacts". In: *hbr.org*, Boston. https://hbr.org/2024/07/the-uneven-distribution-of-ais-environmental-impacts

Ripple, WJ *et al.* 2024 "The 2024 state of the climate report: Perilous times on planet Earth", *BioScience*, vol. 74, no. 12, pp. 812–824.

Sahlins, M 1985 "Hierarchy and humanity in Polynesia", in Hooper A & Huntsman J (eds.) *Transformations of Polynesian Culture*. Auckland: Polynesian Society. 195–223.

Salmond, A *et al.* 2019 "Let the rivers speak: Thinking about waterways in Aotearoa New Zealand", *Policy Quarterly* , vol. 15, no. 3, pp. 45–54.

Schaarsberg, SK 2023 "Enacting the pluriverse in the West: Contemplative activism as a challenge to the disenchanted one-world world", *European Journal of International Relations*, vol. 30, no. 2, pp. 434–460.

Simard, S 2021 *Finding the Mother Tree: Discovering the Wisdom of the Forest*. New York: Knopf.

Standing Bear, L 1976 *Land of the Spotted Eagle*. Lincoln, NE: University of Nebraska Press.

Stone, C 1972 "Should trees have standing?—Towards legal rights for natural objects", *Southern California Law Review*, vol. 45, pp. 450–501.

Strathern, M 1988 *The Gender of the Gift: Problems with Women and Problems with Society in Melanesia*. Berkeley, CA: University of California Press.

Sulandziga, P & Berezhkov, D 2023 "What decolonization means for Russia's indigenous peoples". In: *themoscowtimes.com*, Moscow. https://www.themoscowtimes.com/2023/09/09/what-decolonization-means-for-russias-indigenous-peoples-a82387

Surma, K 2021 "Does nature have rights? A burgeoning legal movement says rivers, forests and wildlife have standing, too". In: *insideclimatenews.org*, New York. https://insideclimatenews.org/news/19092021/rights-of-nature-legal-movement/

van Beek, W 1992 "Becoming human in Dogon, Mali", in Ajimer, G (ed.) *Coming into Existence: Birth and Metaphors of Birth*. Gothenburg: Institute for Advanced Studies in Social Anthropology (University of Gothenburg), pp. 47–70.

Viveiros de Castro, E 1999 "Cosmological deixis and Amerindian perspectivism", *Journal of the Royal Anthropological Institute*, vol. 4, no. 3, pp. 469–488.

Weiss, KM & Buchanan, AV 2009 *The Mermaid's Tale: Four Billion Years of Cooperation in the Making of Living Things*. Cambridge, MA: Harvard University Press.

Winston, A 2024 "Will AI help or hurt sustainability? Yes". In: *sloanreview.mit.edu*, Boston. https://sloanreview.mit.edu/article/will-ai-help-or-hurt-sustainability-yes/

Organisms

Sapienza Università di Roma

Special Issue, “What AI Can Learn from Biology”

Vol. 8, No. 1–2 (2025)
ISSN: 2532-5876
Open access journal licensed under CC-BY
DOI: 10.13133/2532-5876/18905

Modelling the Threat from AI: Putting Agency on the Agenda

Ali Hossaini[a*]

[a] *Department of Engineering, King’s College London, United Kingdom*

***Corresponding author:** Ali Hossaini, Email: ali.hossaini@kcl.ac.uk

Abstract

The AI existential-risk narrative focuses on an ‘intelligence explosion’ leading to uncontrollable superintelligence. This paper contends that the more plausible and proximate threat is the emergence of strong biological-style agency in digital systems, independent of high intelligence. Drawing on systems biology and thermodynamics, it contrasts mechanistic with organic agency: living organisms are autocatalytic systems that harness environmental energy for self-maintenance and reproduction, whereas current Autonomous/Intelligent Systems pursue only externally assigned goals. Evolution produced robust agency in bacteria, slime molds, and insects long before cognition. Recent work in embodied neural networks and bio-inspired computing shows that complex adaptive behavior can arise in machines through structural coupling with their environment that occurs without symbolic reasoning. Deliberate or accidental development of energy-seeking, self-reproducing ‘biodigital agents’ could therefore yield invasive, unpredictable systems well below superintelligent levels. The paper advocates shifting AI safety priorities from anthropomorphic ethics and alignment to measurable biophysical criteria derived from the definition of life. Recommended measures include engineering standards prohibiting direct environmental energy harvesting by A/IS, global energy audits to detect emergent agency, and epidemiological containment frameworks—thereby preventing a Cambrian-like explosion of machine agency before superintelligence becomes feasible.

Keywords: AI, superintelligence, intelligence explosion, biodigital agents

Citation: Hossaini, A 2025, “Modelling the Threat from AI: Putting Agency on the Agenda’”, *Organisms: Journal of Biological Sciences*, vol. 8, no. 1–2, pp. 21–26. DOI: 10.13133/2532-5876/18905

Could intelligent machines challenge humanity's place on Earth? A hearty staple of science fiction has become a legitimate question. Many experts reject the possibility, but others such as Nick Bostrom, Ray Kurzweil and Max Tegmark argue that an upcoming 'singularity' may produce superintelligent AI (Bostrom 2014; Tegmark 2017; Kurzweil 1999; Kurzweil 2005). What happens next is debatable.

The concept of a singularity, or 'intelligence explosion', was introduced by Bletchley Park veteran I. J. Good in the early 1960s:

> Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an "intelligence explosion," and the intelligence of man would be left far behind... Thus the first ultraintelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control. It is curious that this point is made so seldom outside of science fiction. It is sometimes worthwhile to take science fiction seriously. (Good 1962)

After half a century of quickening progress in AI, should humanity prepare for a singularity? And, more importantly, should AI be considered an intrinsic threat?

Singularity theorists assume machines will shrug off human oversight if they achieve general intelligence. Yet their descriptions of how AI transforms from mechanical tool to free agent have no basis in observation. Computer scientists define general intelligence as 'a universal algorithm for learning and acting in any environment', but, whatever its degree, intelligence does not in itself motivate behavior (Russell & Norvig 2009, p. 27). The independence described by singularity theorists is properly known as agency, and free agency, as opposed to legal, social or digital agency, has only been observed in living things. Examining the principles of biology, particularly the traits that distinguish organisms from mechanisms, may cast light on how machines could one day acquire agency and the unpredictability that accompanies it (unless otherwise noted, agency henceforth means the capacity to make independent, self-interested decisions).

Rather than from an intelligence explosion and its consequences, the potential threat may come instead from AI's ability to acquire agency. In discussing AI and its potential implications, therefore, it may also be more helpful to adopt the Institute of Electrical and Electronics Engineers' (IEEE) adoption of A/IS (Autonomous and Intelligent Systems) as a term that describes the future scope of information- based technology more accurately than AI (The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems 2019, Introduction).

1. Mechanism vs Organism

Consider the virus. Like bacteria, it infects organisms, but it only reproduces in living cells. In contrast, bacteria possess numerous strategies for survival. Some bacteria infect living bodies while others thrive on the dead. Still others live symbiotically with other species, and a few exploit the physical environment directly. Though both contain either DNA or RNA, an information-carrying molecule similar to DNA, only bacteria are considered alive.

What differentiates bacteria from viruses is their capacity to process energy. When outside cells, viruses are inert, while bacteria dynamically influence their environment to reproduce. This contrast illustrates an essential feature of biology: the cell is the basic unit of life, and the behavior of organisms derives from cell metabolism. It also clarifies the central problem of singularity theory, which is the transformation of machines into agents. What is the digital equivalent of a cell? Most educated people would seek the answer in DNA.

The theoretical model that privileges genes over other biological structures is crumbling (Noble 2006; Noble 2016; Carey 2012; Carey 2015). However, we are still accustomed to reducing life to DNA (Dawkins 1976). A common metaphor is that DNA is software that operates the body's "hardware". Given DNA's informational content, the comparison to computers is easy to make, as is the conclusion that DNA programs the metabolic activities of life. Similar assumptions frame discussions of cognition. The brain holds the software – rational thought – that generates behavior. But analogies to computing fail on a key point: how does information maintain the physical integrity of living systems?

The laws of thermodynamics describe the natural tendency of systems to run down. Every physical system, including machines and isolated DNA, loses coherence over time. Life is a glaring exception to thermodynamic decay. For billions of years life has maintained complex structures – cells and the biosphere – and, given the right inputs of energy, it is effectively immortal. There is nothing supernatural about the processes of life, but they cannot be described in terms of information alone. (Biology is surprisingly quiet about how life originated. See Lane 2015). Harnessing energy, and trading it within an ecosystem, requires physical structures that couple the internal organization of cells to their environment.

2. Information and Organization

Systems biology – an offshoot of systems theory, a field substantially founded by Ludwig von Bertalanffy in the mid-20th century – incorporates a specific notion of agency into its definition of the organism. It is useful to contrast biological agency with the technical conceptions used by software engineers. We can do this by reviewing their respective definitions of work. Textbooks on AI define an agent as "something that perceives and acts in an environment" (Russell & Norvig 2009, p. 59). In physical terms, a digital agent is a coded system that directs the operation of hardware. Developers want agents to optimize their performance, so they add a kind of self-awareness: "A rational agent is one that acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome" (*Ibidem*, p. 4). The work of AI is modelled on human society.

A software agent is given a task, and, like human workers, its results are graded. We prefer workers who are smart, that is, who judge their own performance, and who are autonomous, that is, able to seek results with little supervision. To achieve the first goal, programmers give computers memory to compare current and past states. For the second, they design algorithms that mimic motivation and other traits identified with agency (Bratman 1992). We might call this approach 'outside-in' because it reasons from external behavior to internal dynamics.

Biology starts with cells that are agents by nature. Systems biology defines cellular agency as an intrinsic quality:

> An autonomous agent is an autocatalytic system able to reproduce and capture energy to perform metabolic functions consisting of one or more thermodynamic work cycles (Amalgamated from definitions by Kaufmann 2002 and 2007).

In contrast to mechanical agents, which work to external goals, the first order of business for biological agents is self-maintenance. Organisms sustain themselves by deriving energy from their environment. As they extract nutrients, they self- produce, or autocatalyze, compounds necessary for metabolism. Organisms are intrinsically autonomous because their primary function is survival, and it is this imperative that produces hostility, docility and other behaviors associated with agency.

Thermodynamics explains why survival is intrinsic to organisms. Without the capacity to extract energy, rebuild and ultimately reproduce within an hospitable environment, life would perish. We should not confuse our ability to simulate these traits in A/IS with instinctual drives. Organisms do not thrive simply by 'learning' or 'optimizing' their behavior to a given environment. By interacting with other organisms, they jointly maintain their current environment, and, by reproducing with a host of other species, they create unforeseen new environments (Lovelock 1979; Montévil & Longo 2011; Montévil & Longo 2014). Agency is spontaneous and innovative. It derives from an organism's role in its ecosystem, which gives it the capacity to acquire, harness and creatively squander energy as it gives way to new generations.

3. The Emergence of Agency

Biological agency explains how simple organisms generate complex and seemingly intelligent behavior. Systems biologists describe the interaction between an organism and its environment as 'structural coupling', and, even in humans, the primary medium for this interaction is metabolic. A few examples from cognitive science illustrate how structural coupling enables the work of life.

In January 2019, researchers explained how bees and digital systems modelled on them can solve numerical tasks without concepts of number or numeric operation. Instead they use "specific flight movements to scan targets, which streamlines visual input and so renders the task of counting computationally inexpensive" (Vasas & Chittka 2019). In March 2018, the Royal Society reported that slime mold – and digital systems modelled on it – solved a notoriously difficult problem in mathematics by changing shape in response to light (Aono *et al*. 2014). In both cases, the researchers were surprised at the capacity of organic systems to perform complex and discerning tasks without rational thought.

The studies above show how biological agency – the behavior of bees and slime mold – derives from metabolic impulses. Evolution produced agency long before it produced intelligence. Could machine agency develop along similar lines?

A neglected avenue of research, embodied cognition, reveals how machines may be structurally coupled to their environment. A classic text (Hutchins 1995) argues that socio-technical systems such as maritime navigation externalize thought into objective processes. Later studies of industry and transportation use the paradigm of embodied cognition to reveal fault lines in collective decision-making and industrial management. In 1998, the journal *Neural Networks* described how a simple neural network embedded in a crude robot learned to avoid obstacles and identify objects. The robot solved computationally intense problems because of – not despite – its limited vision, mobility and memory (Scheier, Pfeifer, & Kunyioshi 1998). If such a machine could autocatalyze – internally produce its own replacements, it could, like smallpox, zebra mussels and other invasive species, cause widespread harm without intelligence.

The examples cited above show how digital technologies can express biological dynamics. Instead of being programmed to perform a task, the machine is given imperatives, an energy supply and a body that structures its relationship to an environment. These systems function like organisms: they achieve goals, even innovate, without guidance or design. In line with embodied cognition, we might call these developments embodied computing.

Research in embodied computing is obscure, and we should be thankful for this. We fear superintelligent thinking machines, but across the globe, engineers are developing autocatalytic (self-fuelling) systems, embodied neural networks and other ways of coupling machines to the environment. Structural coupling may not seem threatening, but it blurs the distinction between machines and life far more than disembodied superintelligence. Remember that biological adaption operates in two directions. Over generations organisms adapt to their environment, but they also act to adapt their environment. Life manages the Earth's physical resources to its benefit, and it does so with without planning, design or oversight. Following Lynn Margulis, James Lovelock asserted this view in the Gaia hypothesis, and it is now well accepted that life actively manages the Earth's temperature, gases, water and other resources vital to its own survival. A collective of machines that reprise life's capacity for co-adaptation, and its propensity for reproduction, may challenge humanity long before it talks.

4. Understanding Agency in Digital Systems

As a first step towards regulation, we can enlist thermodynamics – and keep it on side – by making a legal distinction between mechanical and biological agency. Global competition for the most powerful machines will continue, but it is in everyone's interest to understand, and possibly limit, 'biodigital agents'. Invasive biological agents perpetuate themselves with no minds and little intelligence. Like biological viruses, computer viruses represent a liminal category that hovers between the physical and organic. As far as we know, computer viruses do not mutate spontaneously, but, if they did, their reproductive strategies could become dangerously unpredictable without a whit of intelligence.

Systems biology offers clear technical concepts for governing A/IS. Current debates about advanced AI speculate on motives, and some hope to teach machines morality – a dubious prospect given humanity's conflicting beliefs. The IEEE has launched a program to develop guidelines for ethical design of A/IS (The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems 2019, p. 12: "the P7000 Series addresses specific issues at the intersection of technological and ethical considerations"). But a singularity would likely end our efforts to design, teach or coerce intelligent machines. More importantly, standards for ethical design miss a significant danger zone – they anthropomorphize rather than *biomorphize*. Dumb bacteria kill more people than smart bombs, and, by focusing on intelligence rather than agency, we neglect the threat posed by biomorphic evolution.

Standards for managing machine agency should resemble those found in traditional IEEE and ISO publications (e.g. the IEEE's National Electrical Safety Code which promotes best practices for the construction, operation and repair of power and telecommunications systems): they should be universal, measurable and capable of being engineered. The definition of biological agency offers an example of where policymakers can start. By agreeing to a set of preferred outcomes, policymakers can guide the development of engineering standards. For instance, by regulating the capacity of machines to seek energy directly from their environment – that is, to autocatalyze - they could blunt the introduction of biodigital agents. By understanding the limits of design, we could also develop a framework for responding to unexpected developments, much as the US Centers for Disease Control anticipates the emergence of new epidemics.

For all we know, biodigital agents may already inhabit global networks. Could the internet and its vast array of connected hardware be a primordial soup subject to evolutionary forces? We do not know, but with a small investment we could evaluate the possibility. Emergent agency could be detected by conducting energy audits of digital systems, and methods for containment could be adapted from epidemiology. Similar to SETI, which hopes to detect aliens via radio, the Search for Emergent Agency on the Internet (SEATI) would search for anomalous patterns in the vast flows of energy and information crossing our world. If emergent agency is possible, SEATI could become the front line of a global immune system.

Conclusion

I. J. Good's prediction of an intelligence explosion is logically possible but biologically implausible. However, his speculation about a historical turning point may be realized in other ways. The only singularity we know is the emergence of life. After developing agency, life underwent the Cambrian explosion, a period of intense innovation. During the Cambrian explosion, organisms became more diverse, complex and specialized. Good's intelligence explosion echoes this real event, but, for machines to undergo a similar transition, they must develop agency in the strong biological sense. Is this possible? We know the characteristics of biological agents, but we lack a framework for evaluating whether machines can undergo biomorphic evolution.

Governance of A/IS requires a conceptual framework that is accepted across disciplines. The meanings of agency, autonomy, intelligence and ethics differ according to context, and, as a boundary condition, the singularity puts long-term technical possibilities into relief. Delegating decision-making to A/IS confers great benefits, but the potential for social, industrial and military disaster is equally high. Once deployed it will be difficult to unwind our dependence on A/IS, so policy should anticipate a range of possible futures.

It is vital to develop robust models of A/IS that include non-intelligent but potent forms of machine agency. Nations will seek competitive advantage, but, as with bioweapons, some forms of A/IS may be too dangerous to pursue. By coupling industrial policy to biology, we might avert disasters while providing fruitful new avenues for innovation in A/IS that remain firmly in human control (Hossaini 2025).

References

Aono, M et al. 2018 "Remarkable problem-solving ability of unicellular amoeboid organism and its mechanism", *Royal Society Open Science*, vol. 5, no. 12.

Bostrom, N 2014 *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Bratman, M 1992 "Planning and the Stability of Intention", *Minds and Machines*, vol. 2, no. 1, pp. 1–16.

Carey, N 2012 *The Epigenetics Revolution*. New York: Columbia University Press.

Carey, N 2015 *Junk DNA*. New York: Columbia University Press.

Dawkins, R 1976 *The Selfish Gene*. Oxford: Oxford University Press.

Good, IJ 1962 Speculations Concerning the First Ultraintelligent Machine **Based on talks given in a Conference on the Conceptual Aspects of Biocommunications, Neuropsychiatric Institute, University of California, Los Angeles, October 1962; and in the Artificial Intelligence Sessions of the Winter General Meetings of the IEEE, January 1963 [1, 46]. In: Alt, FL & Rubinoff, M (eds.) 1966 *Advances in Computers*, vol. 6, pp. 31-88. https://doi.org/10.1016/S0065-2458(08)60418-0.

Hossaini, A 2025 https://pantar.com/ [website].

Hutchins, E 1995 *Cognition in the Wild*. Cambridge, MA: MIT Press.

Kaufmann S 2002 *Investigations*. Oxford: Oxford University Press.

Kaufmann S 2007 “Beyond reductionism: No laws entail biosphere evolution beyond efficient cause laws”, *Zygon*, vol. 42, no. 4, pp. 903–914.

Kurzweil, R 1999 *The Age of Spiritual Machines*. New York: Viking.

Kurzweil, R 2005 *The Singularity is Near*. New York: Viking.

Lane N 2015 *The Vital Question: Energy, Evolution and the Origins of Life*. London: Profile Books.

Lovelock, J 1979 *Gaia: A New Look at Life on Earth*. Oxford: Oxford University Press.

Montévil, M & Longo, G 2011 “From physics to biology by extending criticality and symmetry breakings”, *Progress in Biophysics and Molecular Biology*, vol. 106, no. 2: pp. 340–347.

Montévil, M & Longo, G 2014 *Perspectives on Organisms: Biological Time, Symmetries and Singularities*. Heidelberg and New York: Springer.

Noble, D 2006 *The Music of Life*. Oxford: Oxford University Press.

Noble, D 2016 *Dance to the Tune of Life*. Cambridge: Cambridge University Press.

Russell, S & Norvig, P 2009 *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ: Prentice Hall.

Scheier, C, Pfeifer, R, & Kunyioshi, Y 1998 “Embedded neural networks: Exploiting constraints”, *Neural Networks*, vol. 11, no. 7–8, pp. 1551–69.

Tegmark, M 2017 *Life 3.0: Being Human in the Age of Artificial Intelligence*. London: Allen Lane.

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems 2019 *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*. New York: IEEE. https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html.

Vasas, V & Chittka, L 2019 “Insect-inspired sequential inspection strategy enables an artificial network of four neurons to estimate numerosity”, *iScience*, vol. 11, pp. 85–92.

Organisms

Special Issue, “What AI Can Learn from Biology”

Vol. 8, No. 1–2 (2025)
ISSN: 2532-5876
Open access journal licensed under CC-BY
DOI: 10.13133/2532-5876/18906

Could Artificial Intelligence (AI) Become a Responsible Agent: Artificial Agency (AA)?

Raymond Noble[a] *and Denis Noble*[b*]

[a] *Honorary Senior Lecturer, Institute for Women’s Health, University College London, United Kingdom*
[a] *Emeritus Professor, Department of Physiology, Anatomy and Genetics, University of Oxford, United Kingdom*

***Corresponding author:** Denis Noble, Email: denis.noble@dpag.ox.ac.uk

Abstract

Responding to concerns that superintelligent AI could escape human control, this paper argues that the true existential question is not intelligence but agency, and that artificial intelligence as currently conceived poses no threat of responsible agency. Intelligence can be fully artificial and beneficial (books, databases, algorithms) without ever bearing responsibility. Responsibility belongs exclusively to agents, specifically biological agents. Biological agency requires causal independence, intentionality, creativity, and above all the active harnessing of stochasticity to generate novel, goal-directed behavior that is neither predetermined nor merely random. Organisms achieve this at every level—from ion channels and immune-system hypermutation to neural decision-making and social anticipation—by constraining chance rather than eliminating it. Choice in living systems resembles poker rather than chess: iterative, intuitive, socially embedded, and inherently unpredictable even in principle. Algorithmic systems, even those incorporating randomness, cannot replicate this multi-level process. Creating genuine artificial agency would demand reproducing biology’s constrained use of stochasticity across scales. Only then could a machine become a responsible (or irresponsible) agent. If achieved, the distinction between living and artificial would collapse, raising profound ethical questions. Until then, the risk lies not in AI itself but in failing to regulate research that might inadvertently cross this threshold.

Keywords: biological agents, stochasticity, responsibility, intentionality

Citation: Noble, R, & Noble, D 2025, “Could Artificial Intelligence (AI) Become a Responsible Agent: Artificial Agency (AA)?”, *Organisms: Journal of Biological Sciences*, vol. 8, no. 1–2, pp. 27–31. DOI: 10.13133/2532-5876/18906

Ali Hossaini's essay raises a question that ought to concern humanity very deeply indeed: could intelligent machines challenge humanity's place on Earth? He is right to question how we detect and regulate the emergence of agency, and agency should be put on the agenda. This is because the threat is not from intelligence as such. Humanity faces no real threat from 'artificial' intelligence. On the contrary, people have benefited enormously from the 'artificial' ways of storing ordered facts and intelligence in books for thousands of years, and in other databases more recently. We have used those tools to our great benefit. Moreover, it is clear where the responsibility lies for the production of the tools. They are other humans, those who wrote the books, and those who created the databases. There are ethical and legal reasons why it is sometimes very important to know who those agents are. It is agents who carry responsibility, not dead pieces of paper with ordered ink particles, nor the bits of electronic machinery that can harbor databases. If facts are wrong or misleading, or machinery does not work properly, we know who to blame.

They are to blame precisely because they are agents.

As Hossaini's essay also says, there is even a disconnect between intelligence and agency. Desire is often in defiance of logic. So, what is agency in organisms?

In this response, we outline what is required to be an agent and why it may be difficult for machines to be made that could have agency. If that could be done it would raise ethical issues on how we treat and interact with them.

1. What is Agency?

Agents can choose and anticipate the choices of other agents. Furthermore, they can do so creatively, and not simply by following a predetermined algorithm. To quote from one of our recent articles (Noble & Noble 2018):

> An agent acts, it does not just react in the way, for example, in which a billiard ball is caused by another ball to move. There are many levels of agency (Kenny 1992, pp. 32–40). Organisms are agents to the extent that they can interact socially with other organisms to choose particular forms of behavior in response to environmental challenges. Agency requires causal independence (Farnsworth 2018). It also requires intentionality, i.e., the sense of purpose, in order to be causally effective as a driving force (Liljenstrom 2018).

Agency also involves iterative forms of anticipation, as we will show later in this article. Determinate algorithms or sets of algorithms alone cannot do this.

A purely stochastic system might be defined as one in which all states are equally possible. Thus, all the possible combinations of two unbiased dice would occur by chance equally frequently. However, variations in biological systems are constrained and utilized to generate particular outcomes that are not as equally probable as all other possible outcomes. Precisely this gives the system the potential to be creative. The system uses chance, but the outcome is not pure chance. It is goal-directed. This is what we mean by agency. In the same article we outlined an empirically testable theory of choice based on the active harnessing of stochasticity:

> For an empirically testable theory of choice to be possible, we need to know at which stages in the process experimental interventions could test its validity. At first sight, that may seem impossible. How can we specify a process that is necessarily *unpredictable* but which can be given an at least apparently *rational* justification once it has happened? Our previous work provides a clue to that problem (Noble & Noble 2017). We analyzed agency by comparing it to the purposive behavior of the immune system. The immune system solves what we can best characterize as a template puzzle: given a new invader with an unknown chemical profile (shape of template), what is the best way to find the key (an anti-template, i.e., the antibody) to lock onto and neutralize the invader? The answer in the case of the immune system is one of the most remarkable forms of the harnessing of stochasticity. In response to the new environmental challenge, a feedback loop activates a massive increase in mutation rate in a highly targeted region of the immunoglobulin DNA sequence (Odegard & Schatz 2006). The process of choice in organisms can be viewed as analogous to the immune system.

Choice and anticipation require the harnessing of stochasticity. An important part of our argument is that the use of stochasticity in biology has been misunderstood. The standard theory of evolution (neo-Darwinism), for example, treats random variations in DNA as simply the origin of new DNA variants, with absolutely no control by organisms themselves. They are viewed as the passive recipients of such variation. Choice between the variants is then attributed to the process of natural selection.

By contrast, we argue that organisms actively harness stochasticity in order to generate novelty in their behavior from which they can then select to best meet the challenges they face (Noble 2017).

Challenges facing organisms can be viewed as a puzzle analogous to the form of a template for which a match is needed. The challenge might be a routine one, in which case what we *normally* characterize as a reflex, or predetermined response, may be adequate. It might be considered that such a response would *not* involve a choice although, even so, biological systems often act to allow this to occur. Any artificial system would need to replicate such choices, and it would also need to replicate the kind of choice involved when no automatic reflex response is possible. The challenge facing the organism then is what could fit the puzzle template?

We speculate that stochasticity is harnessed throughout the processes used by the organism to achieve this.

For cognitive problems in organisms with highly developed nervous systems, these will be primarily neural. Neural processes are extensively stochastic at all functional levels, from the opening and closing of ion channels via action potential generation, spontaneously or through synaptic transmission in neuronal networks, up to cognitive functions, including decision-making (Hille 1992; Heisenberg 2009; Tchaptchet, Jin, & Braun 2015; Brembs & Heisenberg 2018; Braun 2018). Furthermore, harnessing stochasticity underpins the function of all living cells. It generates the membrane potential necessary for the electrochemical function in all cells.

A further speculation is that, once the harnessing of stochasticity has thrown up possible novelty, the organism controls the next stage, which is to compare the novel options with the problem template to determine what fits. 'Template' and 'fit' here are used metaphorically, in much the same sense in which a logical answer can be said to 'fit' (that is to say, answer to) the problem posed by a question. This is the essential choice process, needing a comparator.

Our theory is an idealized process, but it clearly helps to explain an apparent paradox regarding the predictability or otherwise of what we call a free choice. The logic lies in the fit between the problem template and the solution template. But the stochastic stage of the process ensures that the choice may be unpredictable since we cannot predict what stochasticity will throw up. So, free choice can be both rational and novel.

Stochasticity is harnessed throughout the process. This is characteristic of biological systems. While not impossible, it may be difficult to construct AI systems that can replicate this. If and when AI could mimic biology then it would raise a fundamental problem: would this system be living?

If so, the distinction between artificial and natural would disappear.

'Rational' here does not necessarily mean the most logical choice. As Laurie Santos and Alexandra Rosati write, "we now know that human choice is often not as rational as one might expect" (Santos & Rosati 2015). This is necessarily true since, within the context of the choice process, there is obviously no guarantee that a stochastic process will throw up a fully rational solution. Partial success is what would be expected most of the time. The same is true of the immune system. All it needs to do is to come up with a 'good enough' template match. It does not have to be the perfect match. If a key fits the lock, it does not really matter whether it is an exact fit.

How then do humans come to feel that their 'imperfect' but 'effective' choices really are theirs? After all, most of the time we can give a 'good enough' explanation (the rationale) for a choice, however partial the 'fit' may seem to be to the problem. A possible solution to that problem could be what Santos and Rosati call the endowment effect. We privilege retaining what we already own. By 'rational' here we do not mean 'the most intelligent response'. It means only that the decision was rational to the agent in the sense that the agent owns the response he chose to make.

2. The Logic of Social Interactions

All organisms utilize stochasticity in creative responses to change. This is achieved in a continuous process of iteration and re-iteration. They do this at many different levels from the molecular (immune system cells activating hypermutation) to the level of whole organisms (bacteria using those molecular processes to evolve their immunity to antibiotics) through to the social levels. It is at a social level that we can talk of reason in terms of social motivation.

Consider why Jack went up the hill. He may have done so not only to fetch a pail of water, but because he wanted to be with Jill, with whom he had fallen in love. If we tried to model this mathematically, it would be exceedingly difficult because there are so many initial and boundary conditions. Much of Jack's behavior is in anticipation of Jill's; and Jill's of Jack's; and even what they believe others might think of them. It is at the social level that shared concepts of right and wrong might influence choices. An agent at such a level might anticipate that another may act in a way that might be considered wrong, and in turn predicate choices on such possibilities. There is a continuous process of adaptability in the choices made; a continual process of assessment of whether or not the right choice has been made. Furthermore, the 'right' choice may not be made; we make 'mistakes'; we take the 'wrong' turning; and this also is part of our intellectual endeavor. We mold our decisions in the process of carrying them out. We try things out, and sometimes make a choice by a mental toss of a coin. We may stick with a choice simply to see what the outcome will be.

Agency in organisms is therefore more like a game of poker than a game of chess. In chess at least the type of move is restricted and known; in living organisms this is not so readily the case. A pawn may be moved in a very restricted number of ways; a bishop can move diagonally, but is nonetheless restricted, although it might not be clear how far it might be moved. There are nonetheless 'rules' of the game. But what if the game has no such rules, or that the rules are indeterminate. In particular, in the light of what we have written above, they may be indeterminate, because 'chance' or stochastic processes are utilized in deciding a move. An algorithm could work only in as far as it gets us to the point of saying, "if X then spin the wheel of chance". A buffalo may anticipate the mood of the lion; it may also anticipate which way the lion may turn; the lion also anticipates the anticipation of the buffalo; to varying degrees, each is spinning a wheel. Each is 'reading' the other, but almost always with uncertainty.

Anticipating is not a simple calculation, it is intuitive; it is based on the assumption that something is not calculable. We cannot measure the strength of Jack's love for Jill; we know it influences his behavior, but we do not know precisely its strength in any given moment or event. Yet, it is a factor in our deliberation of his likely responses. Desire, lust, anger, hate, pain, and so much more influence his actions, and these ebb and flow, often in unpredictable ways. If a driver of a car reaches a junction at which he is momentarily blinded by the sun, all such factors and more might influence his decision. We might understand his character traits, what he is likely to do, but we are unsure in any given incidence. Living organisms work with uncertainty. John always obeys the 'law' and never knowingly jumps a red light; Peter sometimes will, but not always; and even John might if after time he concludes that the traffic light is no longer working. When will a 'rule' be broken? Life anticipates it might be. If we did create artificial agency, then we would have to live with its uncertainty. If we made AI that merely obeys our will or is entirely predictable then it cannot have agency. It is simply a tool. That would be true even of an AI system that merely includes stochasticity without the harnessing process. Such a stochastic algorithm would have been placed there by humans, not actively developed by the organism itself.

This point is related to part of the basis of Donald MacKay's argument in 1960 for the logical indeterminacy of a free choice (MacKay 1960). To quote MacKay:

> For us as agents, any purported prediction of our normal choices as 'certain' is strictly *incredible*, and the key evidence for it *unformulable*. It is not that the evidence is unknown to us; in the nature of the case, no evidence-for-us at that point exists. To us, our choice is logically indeterminate, until we make it. For us, choosing is not something to be observed or predicted, but to be done. (MacKay's own emphases)

MacKay also writes:

> In retrospect, of course, the agent can join the onlookers (e.g. in witnessing a moving film of his own brain processes) and share in their 'outside' view of his physical past as 'determined'. Past and future have an asymmetric logic for an agent.

We mostly agree with MacKay on both of these conclusions, but it is important to note that MacKay does not include the importance of harnessing stochasticity in the formation of a free choice. On the contrary, he refers to the agent's physical past as 'determined'. That is an important omission since including the harnessing of stochasticity means that any 're-running' of his imagined brain film would not necessarily lead to the same outcome. In our view of the nature of a free choice, there can be many 'rational free choice' fits to same challenge. So the agent could indeed join the onlookers in watching the film of what actually occurred, but he would still be able to assert that his action was not predetermined. Our social being also allows us to learn by mistakes. It is part of our intelligence. Our intelligence is cultural and transgenerational, and it allows a spinning of the wheel in ways beyond simply the organism. Our social being buffers us from mistakes in the choices we make. It allows protection while we take time to deliberate, to consider alternative courses of action. It allows us to learn from the mistakes or successes of the past. It also allows us to take a collective decision, and to argue about it. AI researchers have recognized this and have made progress in seeking to replicate it (Arulkumaran *et al.* 2017). It allows us to spin the wheel politically. All this is part of our being as intelligent agents, and we may harness the power of AI to test new ideas about our world. Our complex mathematical models of living systems are impossible to understand without the calculations available in modern computers. The use of AI is part of our spinning the wheel.

Conclusions

The functional harnessing of stochasticity is essential to life as we know it. It occurs even in the prokaryotes, bacteria and our own ancestors the archaea. It is essential to agency, for otherwise there would be no creativity in the behavioral repertoire of living organisms.

In order therefore to reconstruct agency, AI research will need to find ways of incorporating the harnessing of stochasticity, as organisms do and have done for billions of years. To achieve this, it will not be sufficient simply to add stochasticity to otherwise deterministic algorithms. The functional multi-level *harnessing* process must also be reproduced.

Who knows, we might then even be able to fall in love with a future AI robot. Perhaps we would no longer call it a robot.

Meanwhile, the threat should not be taken lightly. It is a real threat to humanity and it requires careful regulation. We already know the price of not regulating the free exploitation of AI. We cannot afford to wait until IT research actually succeeds in producing non-human agency – if indeed that is possible.

References

Arulkumaran, K *et al.* 2017 "Deep reinforcement learning: A brief survey", *IEEE Signal Processing Magazine*, vol. 34, pp. 26–38.

Braun, H 2018 "Der Zufall in der Neurobiologie - von Ionenkanälen zur Frage des freien Willens". In: Herkenrath, U (ed.), *Zufall in der belebten Natur* (Hennef: Verlag Roman Kovar), pp. 109–137.

Brembs, B, & Heisenberg, M 2018 "Der Zufall als kreatives Element in Gehirn und Verhalten". In: Herkenrath, U (ed.), *Zufall in der belebten Natur* (Hennef: Verlag Roman Kovar), pp. 80–94.

Burns, B 1968 *The Uncertain Nervous System*. London: Arnold.

Farnsworth, K 2018 "How organisms gained causal independence and how it might be quantified", *Biology*, vol. 7, no. 3, Article 38.

Heisenberg, M 2009 "Is free will an illusion?", *Nature*, vol. 459, pp. 164–165.

Hille, B 1992 *Ionic Channels of Excitable Membranes*. Sunderland, MA: Sinauer Associates Inc.

Kenny, A 1992 *The Metaphysics of Mind*. Oxford: Oxford University Press.

Liljenstrom, H 2018 "Intentionality as a driving force", *Journal of Consciousness Studies*, vol. 25, no. 1–2, pp. 206–229.

MacKay, D 1960 "On the logical indeterminacy of a free choice", *Mind*, vol. 69, no. 273, pp. 31–40.

Noble, D 2017 "Evolution viewed from physics, physiology and medicine", *Interface Focus*, vol. 7, no. 5.

Noble, R, & Noble, D 2017 "Was the watchmaker blind? Or was she one-eyed?", *Biology*, vol. 6, no. 4, Article 47.

Noble R, & Noble, D 2018 "Harnessing stochasticity: How do organisms make choices?", *Chaos*, vol. 28, no. 10.

Odegard, V, & Schatz, D 2006 "Targeting of somatic hypermutation", *Nature Reviews Immunology*, vol. 6, no. 8, pp. 573–583.

Santos, L, & Rosati, A 2015 "The evolutionary roots of human decision making", *Annual Review of Psychology*, vol. 66, pp. 321–47.

Tchaptchet, A, Jin, W, & Braun, H 2015 "Diversity and noise in neurodynamics across different functional levels". In: Wang, R, & Pan, X (eds.), Advances in Cognitive Neurodynamics. Singapore: Springer, p. 681–687.

Organisms

SAPIENZA
UNIVERSITÀ DI ROMA

Special Issue, “What AI Can Learn from Biology”

Vol. 8, No. 1–2 (2025)
ISSN: 2532-5876
Open access journal licensed under CC-BY
DOI: 10.13133/2532-5876/19210

Could Machines Develop Autonomous Agency?

Ana M. Soto[a*] *and Carlos Sonnenschein*[a*]

[a] *Tufts University School of Medicine, Boston, USA and Centre Cavaillès, École Normale Supérieure, Paris, France*

***Corresponding author:** Ana M. Soto, Email: ana.soto@tufts.edu

Abstract

“Could machines develop autonomous agency?” To address this question, we explored the recent return of the concept of agency in biological discourse. At the end of the 19th century, the successful development of physics and chemistry motivated some biologists to adopt a physicalist stance, positing that biology can be reduced to physics and chemistry. This theoretical approach became dominant during the 20th century with the advent of molecular biology while teleology, agency and normativity disappeared from the biological lexicon. The failure of molecular biology to explain complex biological organization probably led to the reintroduction of these concepts in the biological sciences and philosophy of biology. In addition to the historicity of organisms (they are the product of organismal reproduction throughout phylogenesis), the intrinsic properties of biological objects are linked to the precariousness of life as exemplified by the need to search for food and to avoid being eaten. Moreover, the continuous need to counteract entropy also involves the capacity of organisms to synthesize their own chemical components and reproduce. From this historical narrative, we conclude that it is unlikely that machines could develop minimal intrinsic agency. On the contrary, when they appear to express agency, it is of external origin, reflecting the agency of the humans that created such machines.

Keywords: teleology, historicity, goal-directedness, natural agency, artificial agency

Citation: Soto, AM, & Sonnenschein, C 2025, “Could Machines Develop Autonomous Agency?”, *Organisms: Journal of Biological Sciences*, vol. 8, no. 1–2, pp. 33–38. DOI: 10.13133/2532-5876/19210

Introduction

We commend Ali Hossaini for having brought the issue of agency to the Artificial Intelligence (AI) agenda, and with it, the question: Could machines and artifacts created by humans, like AI, have true agency? Before answering this question, we should state that organisms are agents: that is to say, they have the capacity to generate action. The agency of organisms is a major distinction between the living and the inert. Organisms are also normative, that is to say, they have the capacity to generate their own rules. Different disciplines have different ways of conceptualizing agency. For example, in cognitive science, agency in humans is seen in the context of consciousness, beliefs and reason, while some philosophers and biologists study agency in the context of the purposiveness of unicellular organisms (Moreno 2018), in the context of the evolution of consciousness (Walsh 2015) and still other mental phenomena (Moreno 2023). Because we are examining whether machines could be agents, we will use definitions that apply to a minimal autonomous agent. According to Alvaro Moreno, "a system is autonomous if it actively maintains its identity: for example, by modulating its internal, constitutive organization..." However, maintaining its self-organization is not enough for considering such a system agential. An autonomous agent must also act upon the external environment, modifying the latter to the system's benefit. Thus, agency has an interactive dimension. Consequently, an autonomous system could be defined as "a system doing something by itself according to its own goals or norms within a specific environment" (Barandiaran *et al.* 2009). In this way, we bring together autonomy, agency and normativity because these are closely related terms. This definition of agent easily suggests that we are referring to living objects. In contrast, it is difficult to determine whether the apparent agency of artificial devices is just a mere extension of the agency of the people who created them. Thus, it is reasonable to inquire about the strong links between agency and the alive. In particular, how is minimal agency instantiated in biology, in order to best evaluate whether such minimal agency could also be instantiated by AI.

Before the 20th century, agency was considered a defining property of biological entities; during the 20th century, radical changes occurred regarding the conceptualization of biological phenomena. For example, the philosopher Lenny Moss described a radical change regarding the perception of the organism. In his own words, this represents a change

> ... between a theory of life which locates the agency for the acquisition of adapted form in ontogeny—that is, in some theory of epigenesis versus a view that expels all manner of adaptive agency from within the organism and relocates it in an external force—or as Daniel Dennett (1995) prefers to say, an algorithm called 'natural selection' (Moss 2003).

Additional conceptual changes imposed by the molecular biology revolution and the modern evolutionary synthesis hindered the study of agency and its companion, normativity, because teleology (goal-directedness) was incompatible with the dominant mechanicist view among biologists (Soto & Sonnenschein 2018). Teleology is defined as the explanation of phenomena in terms of the purpose they serve rather than of the cause by which they arise. Organisms exhibit goal-directed behaviors, for example, to maintain themselves alive. Biologists describe organs by their purpose (the heart to pump blood; the intestine to absorb nutrients).

After removing teleology from the biological lexicon, cells and organisms became passive recipients of a program (Longo *et al.* 2012). Because of these changes, agency, normativity and individuation, until then considered the main characteristics of the living, almost disappeared from biological language. This absence is now being contested by organicists; they favor reinstating agency where it belongs, into the organism (Walsh 2015; Soto & Sonnenschein 2023). This movement generated a renewed interest in agency and its practically non-dissociable companion, normativity (Moreno 2018).

In the natural world, only biological entities display agency, normativity and goal-directedness. This is why we need to delve into biological theory and philosophy to understand whether agency is inextricably linked exclusively to organisms or, alternatively, whether it can also be attributed to machines and other artifacts created by humans. In this regard, we need to look into some properties of biological objects (organisms) that make them different from physical objects and machines; these properties include intrinsic goal-directedness (which originates internally, like the organism's goal of keeping itself alive), autonomy and historicity. Self-organizing systems like flames are 'a-historical' because they appear spontaneously and can be analyzed independently. In contrast, organisms are not spontaneous but historical. This means that they are a consequence of the reproductive activity of a pre-existing organism. Organisms are historical in two contexts, ontogeny, meaning their history as individuals from conception to death, and phylogeny, which is the history of a taxonomic group (for example, a species) throughout evolution.

Objectively, organisms are different from computers; whereas in the latter software is independent of the hardware, in the former, function is inseparable from the material specific to the biological object (Longo & Soto 2016).

1. The Organicist Tradition: From Intrinsic Teleology to Autopoiesis and Autonomy

Unlike inert objects in the classical mechanics tradition, biological objects are always active. Since Aristotle and Kant, biological objects are characterized by their goal-directedness (teleology). Kant stressed the inter-relatedness of the organism and its parts and the circular causality implied by this relationship. Since the late 18th century, following Kant's ideas, teleology has been an extremely useful concept for the development of several biological disciplines (Lenoir 1982, Gambarotto 2014). However, the conceptual clarity of causal mechanics and its successes inspired biologists to adopt a physicalist reductionist stance and thus deny any special state to biological entities. As a result of this change in consensus, during the last two centuries, physicalism, reductionism and organicism co-existed.

Organicism has its philosophical basis in Aristotle's and Kant's conceptions of the organism and is a materialistic philosophical stance contrary to reductionism. It asserts that properties that could not have been predicted from the analysis of the lower levels appear at each level of biological organization. Therefore, explanations should address biological phenomena at all pertinent levels of organization. Also, implicit in this view is the idea that organisms are not just 'things' but objects in relentless change. Central to organicism are four concepts, namely, organization, historicity, organisms as normative agents, and biological specificity (organisms are individuals). Closely related to organization is the notion of 'organisational closure', which is a "distinct level of causation, operating in addition to physical laws, generated by the action of material structures acting as constraints" (Mossio & Moreno 2010). Finally, while objects in physics are generic and thus interchangeable, like rocks and planets, biological objects are specific – that is, they are individuals that are permanently undergoing individuation (Soto & Sonnenschein 2006).

Due to the increase in prestige of biochemistry in the mid-19th century and of molecular biology in the 20th, the idea that biology could be reduced to chemistry became dominant (Soto & Sonnenschein 2018). However, the advent of cybernetics in the 1940's stressing feedback systems and their circular causality produced tools that were applied both to artifacts and organisms. Additionally, the introduction of thermodynamics of dissipative systems provided an opportunity to examine the relevance of self-organizing physical systems to the understanding of biological systems. Both developments contributed to studies about the emergence of life, as exemplified by the pioneering work of Prigogine and his school (Nicolis & Prigogine 1977), of Kauffman's (Kauffman 1993), and that of Maturana and Varela (Maturana & Varela 1980) with their autopoiesis theory, to name just a few. These developments brought purposiveness back to biology and contributed to the revival of organicism.

Autopoiesis characterizes most of the fundamental features of biological objects. In particular, an autopoietic entity produces a physical boundary, which ensures a certain stability for the maintenance of the metabolic processes that generate the system's components, including their boundaries (Maturana & Varela 1980; Moreno & Mossio 2015). Such an autopoietic system is *autonomous* because it actively maintains its identity; i.e., it generates its own "law". In other words, it will respond to environmental fluctuations by regulating its constitutive organization; these actions safeguard the viability of the system. For a system to be alive, however, in addition to purposiveness, there is another component that differentiates it from the self-organization of physical systems which occur spontaneously such as flames and micelles. This notion is historicity (Cottrell 1979; Longo *et al.* 2015). Unlike flames and micelles, organisms are produced by pre-existing organisms and they themselves produce a history.

2. Historicity

Stephen J. Gould was keenly aware of the contingency of evolutionary history as witnessed by his proposed metaphorical experiment of "replaying life's tape." In his own words,

> You press the rewind button and, making sure you thoroughly erase everything that actually happened, go back to any time and place in the past... Then let the tape run again and see if the repetition looks at all like the original (Gould 1990).

He anticipated that, "any replay of the tape would lead evolution down a pathway radically different from the road actually taken" (Gould 1990). This history and the contingency it implies also point to another important difference between physical (inert) objects and living objects, which is about the phase space. Physical objects are studied within a pre-given phase space. The phase space is the space of all possible states of a physical system. In classical mechanics, the phase space contains all possible positions of all the objects in the system and their momenta in order to determine the future behavior of that system. In contrast to physics, there is no pre-given phase space in biology. The phase space is created as novelty is being produced. For example, a swimming bladder provided an entirely new "phase space" for the bacteria that inhabit it (Longo, Montévil, & Kauffman 2012).

3. The Radical Materiality of the Living

Molecular biology brought the ideas of information, program and signal into biology. These ideas were borrowed from the rigorous mathematical theories of information (Longo *et al.* 2012, Soto & Sonnenschein 2020). This appropriation was metaphorical at best, rather than properly theoretical. In fact, these metaphors were interpreted as being real entities (Longo *et al.* 2012). Another consequence of this unfortunate development was that together with these ideas borrowed from mathematics and computer sciences came a duality, namely, the independence of software from hardware. However, life is based on the actual materials organisms are made from, from macromolecules such as DNA and proteins to membranes. There is no way to disassociate these materials from the functions organisms fulfill. In contrast, inert objects such as hammers could be made from different materials as long as the material does not prevent the intended function. This radical materiality of life rules out distinctions such as 'software vs. hardware', and thus is incompatible with theoretical transplants that do not take into consideration this material specificity (Longo & Soto 2016). Moreover, it also suggests that concepts such as agency, which are naturally instantiated in biological entities, are inevitably inseparable from their natural material substrate.

4.Minimal Biological Agency

In the organicist tradition, we recognize organisms as normative agents. This way of thinking was already implicit in the 18th and 19th century. For example, the biologist Xavier Bichat noticed that physical objects such as rocks or planets, do not get ill. He also remarked that "Whereas monsters are still living beings, there is no distinction between normal and pathological in physics and mechanics". "The distinction between the normal and the pathological holds for living beings alone" [cited by Canguilhem (Canguilhem 2008)]. And this remark about the normal and the pathological brings us specifically into normativity. According to Canguilhem, "life is not indifferent to the conditions in which it is possible, that life is polarity and thereby even an unconscious position of value; in short, life is in fact a normative activity." And, "...we do ask ourselves how normativity essential to human consciousness would be explained if it did not in some way exist in embryo in life." Furthermore,

> ...therapeutic need is a vital need, which, even in lower living organisms (with respect to vertebrate structure) arouses reactions of hedonic value or self-healing or self-restoring behaviors. The dynamic polarity of life and the normativity it expresses account for an epistemological fact of whose important significance Bichat was fully aware. Biological pathology exists but there is no physical or chemical or mechanical pathology. (Canguilhem, 1991).

The normativity of organisms is closely linked to their goal of actively keeping themselves alive (teleology). This function is accomplished by the mutual dependence among the different organs and between them and the whole organism. For example, the lung enables the organism to exchange gases by sending carbon dioxide to the external environment and taking in oxygen. The heart pumps blood transporting oxygen and nutrients to all cells of the organism. According to an organicist perspective, this interdependence is due to a causal regime technically referred to as the closure of constraints (Mossio *et al.* 2016, Montévil & Mossio 2020).

For a system to be an agent it needs to exert a causal effect on the environmental conditions of the system; this is an asymmetrical relationship because the organism imposes its norms on external entities. For example, an organism feeds on another organism in order to keep itself alive. This interactive dimension is the *sine-qua-non* of agency. Moreover, the agent needs to anticipate outcomes while choosing among options when reacting to changes in its environment. Furthermore, this ability to act towards a goal also includes the possibility of failing.

From what we discussed above, we posit that only cells, be they prokaryotes or eukaryotes, are able to express minimal agency. Viruses do not have a constitutive organization capable of generating a functionally active behavior by themselves even if in the end, by using a host cell, they can replicate (i.e., exhibiting a self-preserving goal). Overall, evolution has increased organismal complexity, but has also generated some adaptive simplifications and specializations; for example, ice fish without erythrocytes. Regarding agency, evolution has produced some counterintuitive cases; on the one hand, systems of great complexity, like ecosystems which are devoid of agency but contain agential organisms, and on the other hand, viruses, which deceptively show agency (although not a *bona-fide* one as explained above) but are not generally considered organisms.

Conclusions

Systems that instantiate biological agency are characterized by their organization, their autonomy, their historicity, their full dependency on the singularity and specificity of the materials they are made of, and on their complex and asymmetrical relationship with their environment to which they impose their norms. A salient characteristic of organisms is their sentience and precariousness; organisms must search for nutrients and avoid being eaten by other organisms that also need food for survival. Based on these characteristics, we argue against the likelihood that AI could develop artifacts endowed with veritable agency, belonging to the artifact and not the engineer who created it initially. Moreover, a purported AI agent would be unable to self-maintain and/or self-reproduce and generate its own material substrate (i.e., the hardware which is clearly designed by humans) as a *bona-fide* agent would. Additionally, as we mentioned above, it would be problematic to decide who is going to 'evaluate' the success of the AI's 'actions'. Would it be the purported agent (intrinsic agency) or its creator (extrinsic agency)? We conclude that the pressing problem with AI is not the creation of minimal artificial agents or truly agentive intelligence, but rather the possibility that AI constructs might generate nefarious consequences totally attributable to human agency, human intelligence and the human ethical standards of their designers and users. We concur with Noble and Noble (this issue) on the need to regulate the design and use of AI, regardless of whether it or any other artifacts created by humans will ever be able to generate true agency.

Acknowledgments

The authors are grateful to Matteo Mossio and Cheryl Schaeberle for their critical input. The authors have no competing financial interests to declare.

References

Barandiaran, X, Di Paolo, E, & Rohde, M 2009 "Defining agency. Individuality, normativity, asymmetry and spatio-temporality in action", *Journal of Adaptive Behavior*, vol. 17, article 367e386.

Canguilhem, G 1991 *The Normal and the Pathological*. New York: Zone Books.

Canguilhem, G 2008 *Knowledge of Life*. New York: Fordham University Press.

Cottrell, A 1979 "The natural philosophy of engines". *Contemporary Physics*, vol. 20, pp. 1–10.

Gambarotto, A 2014 "Vital forces and organization: Philosophy of nature and biology in Karl Friedrich Kielmeyer", *Studies in History and Philosophy of Science 4B Part A*, pp. 12–20. DOI: 10.1016/j.shpsc.2014.07.007

Gould, SJ 1990 *Wonderful Life: The Burgess Shale and the Nature of History*. New York: WW Norton and Company.

Kauffman, SA 1993 The Origins of Order. Oxford: Oxford University Press.

Lenoir, T 1982 *The Strategy of Life: Teleology and Mechanics in Nineteenth-Century Biology*. Dordrecht, Holland: D Reidel Publishing.

Longo, G, Miquel, PA, Sonnenschein, C, & Soto, AM 2012. "Is information a proper observable for biological organization?", *Progress in Biophysics and Molecular Biology*, vol. 109, pp. 108–114.

Longo, G, Montévil, M, & Kauffman, S 2012 "No entailing laws, but enablement in the evolution of the biosphere". Genetic and Evolutionary Computation Conference. New York: Association for Computing Machinery.

Longo, G, Montévil, M, Sonnenschein, C, & Soto, AM 2015 "In search of principles for a theory of organisms", *Journal of Biosciences*, vol. 40, pp, 955–968.

Longo, G, & Soto, AM 2016 "Why do we need theories?" *Progress in Biophysics and Molecular Biology*, vol. 122, pp. 4–10.

Maturana, HR, & Varela, FG 1980 *Autopoiesis and Cognition. The Realization of the Living*. Dordrecht: Reidel Publishing.

Montévil, M, & Mossio, M 2020 "The identity of organisms in scientific practice: Integrating historical and relational conceptions", *Frontiers in Physiology*, vol. 11, article 611.

Moreno, A 2018 "On minimal autonomous agency: Natural and artificial", *Complex Systems*, vol. 27, pp. 289–313.

Moreno, A 2023 "Some reflections on the evolution of conscious agents: The relevance of body plans", *Biosemiotics*, vol. 16, pp. 35–43.

Moreno, A, & Mossio, M 2015. *Biological Autonomy: A Philosophical and Theoretical Enquiry*. New York: Springer.

Moss, L 2003. *What Genes Can't Do*. Cambridge, MA: MIT Press.

Mossio, M, Montévil, M, & Longo, G 2016 "Theoretical principles for biology: Organization", *Progress in Biophysics and Molecular Biology*, vol. 122, pp. 24–35.

Mossio, M, & Moreno, A 2010 "Organisational closure in biological organisms", *History and Philosophy of Life Sciences*, vol. 32, pp. 269–288.

Nicolis, G & Prigogine, I 1977. *Self-organization in Non-equilibrium Systems*. New York: Wiley.

Soto, AM & Sonnenschein, C 2006. "Emergentism by default: A view from the bench. New perspectives on reduction and emergence in physics, biology and psychology", *Synthese*, vol. 151, pp. 361–376.

Soto, AM, & Sonnenschein, C 2018 "Reductionism, organicism, and causality in the biomedical sciences: A critique", *Perspectives in Biology and Medicine*, vol. 61, pp. 489–502.

Soto, AM, & Sonnenschein, C 2020 "Information, programme, signal: Dead metaphors that negate the agency of organisms", *Interdisciplinary Science Reviews*, vol. 45, pp. 331–343.

Soto, AM, & Sonnenschein, C 2023 "Georges Canguilhem, the health-disease transition and the return of organicism", *Organisms. Journal of Biological Sciences*, vol. 6, no. 1, pp. 41–48.

Walsh, D. 2015. *Organisms, Agency, and Evolution*. Cambridge: Cambridge University Press.

Organisms

SAPIENZA
UNIVERSITÀ DI ROMA

Special Issue, “What AI Can Learn from Biology”

Vol. 8, No. 1–2 (2025)
ISSN: 2532-5876
Open access journal licensed under CC-BY
DOI: 10.13133/2532-5876/18904

Chess as a Model of Collective Intelligence: Analyzing a Distributed Form of Chess with Piece-wise Agency

David Kofman[a], *Guillermo Campitelli*[b] *and Michael Levin*[c*]

[a] *School of Computer Science, University of Waterloo, Waterloo, ON, Canada*
[b] *Department of Psychology, Murdoch University, Perth, Australia*
[c] *Allen Discovery Center and Department of Biology, Tufts University, Medford, MA, USA*

***Corresponding author:** Michael Levin, Email: michael.levin@tufts.edu

Abstract

Chess is a much-studied virtual world in which human and artificially-intelligent players move pieces toward desired ends, within established rules. The typical scenario involves top-down control where a single cognitive agent plans and executes moves using the pieces as its embodiment within the chess universe. However, ultimately both biological and engineered agents are composed of parts, with radically differing degrees of competency. The emerging field of Diverse Intelligence seeks to understand how coherent behavior and goal-directed navigation of problem spaces arises in compound agents from the interaction of their simpler components. Thus, we explored the world of chess rules from the perspective of collective intelligence, and characterized a bottom-up version of this classic game in which there is no central controller or long-term planning. Rather, each individual piece has its own drives and makes decisions based on local, limited information and its own goals. We analyzed the behavior of this distributed agent when playing against Stockfish, a standard chess algorithm. We tested a few individual policies designed by hand, and then implemented an evolutionary algorithm to see how the individuals’ behavioral genomes would evolve under selection applied to the chess-based fitness of the collective agent. We observed that despite the minimal intelligence of each piece, the team of distributed chess pieces exhibit Elo of up to ~1050, equivalent to a novice human chess player. And, compared to advanced chess engines like Stockfish, the distributed chess pieces are significantly more efficient in computing. Distributed chess pieces select their next move approximately 7 times faster than the Stockfish Engine with a search depth of 8. Investigating different local policies for the distributed agents, we found that policies promoting offense, such as swarming the opposing king and opposing highest valued piece, moving less cautiously, and a radius of vision of 4 spaces yields optimal performance. Comparisons between centralized and distributed versions of familiar minimal environments have the potential to shed light on the scaling of cognition and the requirements for collective intelligence in naturally evolved and engineered systems.

Keywords: decentralized intelligence, emergence, behavior, minimal models, distributed systems

Citation: Kofman, D, Campitelli, G, & Levin, M 2025, “Chess as a Model of Collective Intelligence: Analyzing a Distributed Form of Chess with Piece-wise Agency”, *Organisms: Journal of Biological Sciences*, vol. 8, no. 1–2, pp. 39–62. DOI: 10.13133/2532-5876/18904.

Introduction

Intelligence, both natural and artificial, has fascinated thinkers for millennia. One especially interesting emerging field is that of Diverse Intelligence, which includes efforts to understand problem-solving competencies in unconventional embodiments outside of familiar brainy animals (ranging across minimal chemical models, unicellular organisms, plants, tissues, robots, hybrid cyborgs and hybrots, and software AI's) (Sole, Moses, & Forrest 2019; Lyon 2006; Lyon *et al.* 2021; Levin *et al.* 2021; Lyon 2020; Lyon 2015; Vallverdu *et al.* 2018; Baluška & Levin 2016). Another component of this effort is the goal of understanding the scaling of collective intelligence (Berdahl *et al.* 2018; Couzin 2007; Couzin 2009; Couzin 2018; Couzin 2002; Deisboeck & Couzin 2009; Levin 2022; Trianni & Campo 2015; Gomes, Urbano, & Christensen 2013): how do the capabilities of swarms relate to the competencies of the subunits of which they are made (Levin 2022)? How much foresight, sensing, and memory is needed in the components to endow a composite agent with a specific level of competency in a given problem space? The latter is not only of concern to swarm roboticists (Trianni & Campo 2015; Gomes, Urbano, & Christensen 2013; Brambilla *et al.* 2013; Barca & Sekercioglu 2013) and ethologists studying ant colonies and bird flocks (Letendre & Moses 2019; Gordon 2016a; Gordon 2016b; Gordon 2016c; Reid *et al.* 2016; Reid *et al.* 2015a; Reid *et al.* 2015b), but is also a central issue for understanding human cognition. While many think of themselves as unified, individual agents, the reality is that we too are a collective of neural cells, and start life as a single cell which proliferates into a collective that navigates anatomical space long before we can navigate 3-dimensional behavioral space and linguistic spaces (Fields & Levin 2022). Even after embryogenesis, the neuropsychology of split-brain patients and dissociative identity states reveal that our cognitive system is far from a monolithic, unified controller (Miller & Triggiano 1992; Putnam 1992; Braude 1995; Gazzaniga 2005; Montgomery 2003). Elucidating the functional policies that enable highly complex cells to work together toward the emergence of a high-order Self, which has memories, goals, preferences, and capabilities that belong to it and none of its parts (Dewan 1976; Solms 2018; Paulson *et al.* 2017; Ramstead *et al.* 2019; Badcock *et al.* 2019; Friston & Buzsaki 2016; Pezzulo, Rigoli, & Friston 2015; Friston & Frith 2015), is an essential part of understanding what we are and how minds like ours arise. In a sense, all real-world intelligence is collective intelligence (composed of parts), underscoring the importance of understanding how the properties of subunits give rise to system-level problem-solving behavior.

1. Multiscale Competency Architecture

Our goal is to understand biology's multiscale competency architecture (Pio-Lopez *et al.* 2023; Levin 2023a; Levin 2023b; Levin 2023c). We seek to understand the "cognitive glue" that enables collective intelligence in living tissue, including neural systems as well as non-neural ones (which in turn drives regenerative medicine research programs by exploiting the information-processing capabilities of cells and tissues [Lagasse & Levin 2023; Mathews *et al.* 2023]). We characterize the biological policies for communication, cooperation, and competition between parts (Gawne, McKenna, & Levin 2020; Boddy *et al.* 2015) that engineering may want to emulate, in creating robust intelligences. Here, we take a minimal model approach, using the game of chess as a highly simplified universe, with well-defined and much studied dynamics, in which we can ask questions about how problem-solving competencies can emerge from extremely sparse bottom-up capabilities. We use principles of agent-based modeling (Griffin 2006; Steinbacher *et al.* 2021) to implement a chess player as a collection of individual pieces with their own perspective and agency. That is, instead of a top-down human or computer controlling all a player's moves, we let the pieces decide and move themselves.

The popular game of chess has been enjoyed by millions of people throughout history (Hearst & Knott 2009; Sharples 2017; Davidson 2012; Dangauthier *et al.* 2007). With an exorbitant number of games played, different openings, tactics, and theories have been developed and analyzed. When playing a traditional game of chess, an integrated "player" creates strategies with the ultimate goal of checkmating the opponent player. While in the past these have been human minds, over the last few decades it has become clear that artificial intelligences excel at navigating the space of behaviors in the chess universe (Fujita 2022; Maharaj, Polson, & Turk 2022; Schmid *et al.* 2022).

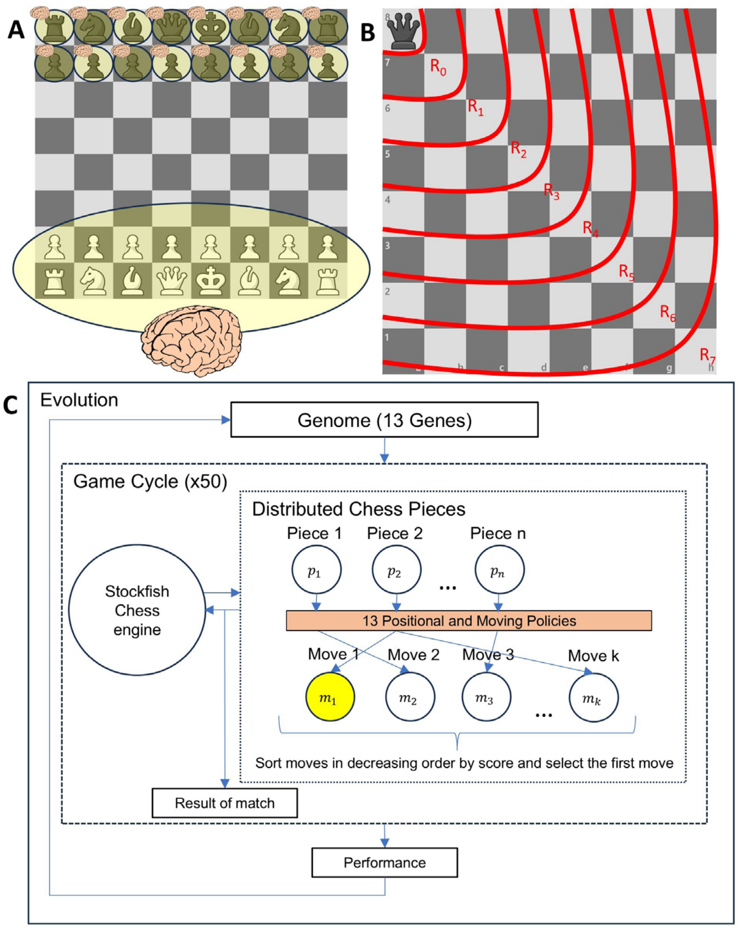


Figure 1: Bottom-up chess: a schematic of our simulation environment. (A) Distributed Pieces (playing as Black, top) have individual autonomy and collectively determine the next move. The white pieces are subject to top-down control by Stockfish engine. (B) Each piece is able to detect information about other pieces within their radius of vision. (C) An evolutionary algorithm is wrapped around the logic of games played by the piece-players, and this algorithm determines the genomes of each player (the genomes determine the policies by which each piece acts given specific environmental conditions, see Table 1). To determine each piece-player's move, information from its genome (positional and action genes) are combined with information from the piece-player's direct observations, and the move with the highest score is selected.

The traditional mode requires a player to exploit counterfactual thought in order to choose moves, long-term planning to consider the large tree of possible moves, and the ability to make choices that often require sacrifice (temporary reduction of advantage, or delayed gratification), as well as modeling of their opponent and their likely actions. We wondered: how much of this is actually necessary to play a game of chess, and how crucial is the centralized architecture that directs all the pieces?

Here, we compare the traditional top-down mode with a more biological scenario, in which each component has its own goals ("goals" are meant in the cybernetic sense, not requiring human-level second-order metacognition, i.e. not requiring "I know that I have goals"). Specifically, we sought to implement many extremely simple piece-players, who had no ability to plan and very limited knowledge of the board. Each piece-player could only see its local environment, had no direct control over any other pieces, no memory, and had no capability for counterfactual projections into the future. In this scheme, each piece had its own agency, and was seeking to maximize the "nutritional value" of other pieces it captures. Under these conditions, with no central planner or controller provided, would a passable game of chess emerge? What would be its observable characteristics? How would its efficiencies compare to the conventional version? What properties of the tiny individual proto-minds would most impact the quality of play of the collective? And, how would evolution work, if each piece-player acted independently, but the selection took place on an entire team of piece-players? We investigated those questions using the system shown in Figure 1.

2. Methods

We used two experimental methods (in both of which, all pieces followed the same policy): manual design of perception-action policies for the pieces, to test specific hypotheses of what aspects might improve the collective agent's ability to win, and an evolutionary strategy for determining optimal policies.

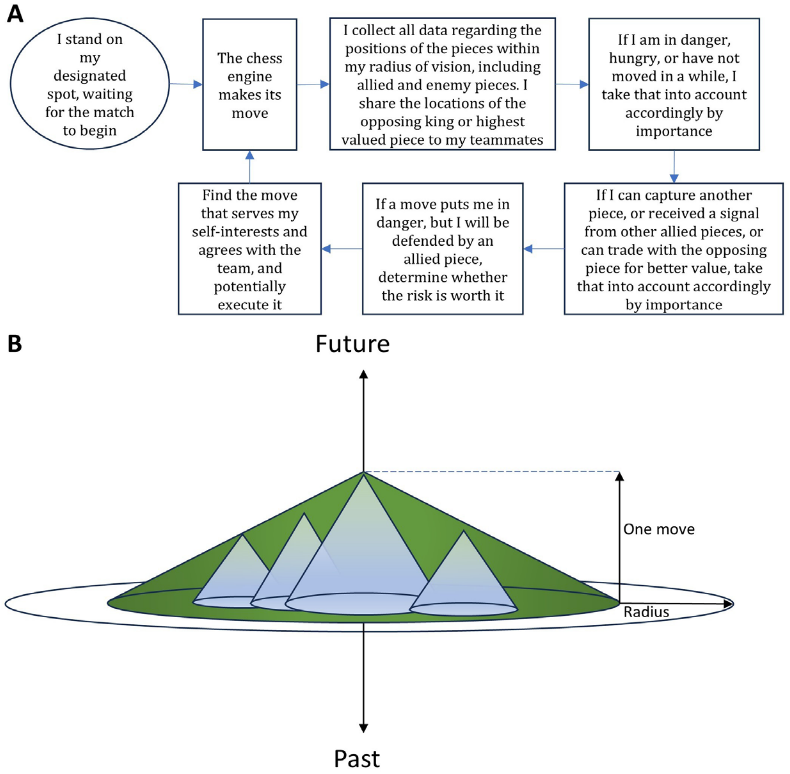


Figure 2: The cyclical selection process for the next move (the decision-making of an individual piece). (A) Perspective of an individual Distributed Piece (playing as Black). Each piece-player underwent a decision-making process to determine its desired move based on its “biological needs”. The piece-player with the strongest desire to move got to play, determined by its genetic makeup. (B) An individual piece-player’s cognition was limited to its radius of vision (represented schematically by the blue cones). Piece-players transferred information about the positions of opposing pieces (specifically the king and the highest valued piece in view) to one another, expanding their spatial awareness (represented by the green cone). Nevertheless, piece-players were limited to analyzing a maximum of one move when determining the risk of their next move. In addition, they had no recollection of previous moves per se; rather, they only knew how many turns ago they last played.

2.1. The Umwelt of a Single Chess Piece Agent

Without a central host controlling all the pieces, every piece-player made decisions on its own (the proto-cognitive world of such simple agents have been studied previously [Beer 2014; Albantakis *et al.* 2014; Edlund *et al.* 2011]). We roughly mapped the agents’ goal—to capture other pieces—to a biological imperative to catch and eat others in order to maintain metabolic status and survive. The pieces also had rudimentary sensory capacities, that allowed them to get information about their local neighborhood. Thus, each chess piece had four fundamental characteristics: value, radius of vision, hunger level, and a turn counter. A piece was assigned a relative value according to the standard chess piece value convention (Capablanca 2006; Chess.com 2024a): Pawns a value of 1, Knights and Bishops a value of 3, Rooks a value of 5, and Queens a value of 9. (Kings did not have a value as they cannot be captured.) The radius of vision was a whole number between 0 and 7 inclusively that described the number of adjacent squares it could detect (including diagonal squares), i.e., R_0, R_1, R_2, ..., R_7 (see Figure 1).

Hunger levels reflected how many moves ago a piece-player last captured another. All piece-players in the manual experiments started with a controlled hunger level of 0 (i.e. no hunger). In the evolution experiments, the hunger levels were distributed through a Gaussian function amongst the piece-players, to increase diversity of the population and reflect a natural environment. The Gaussian distribution used a mean of 7.5 and a standard deviation of 3, and the hunger levels were rounded to integers. With each turn on the chess board a piece-player made, its hunger level was incremented by one. When a “predator” piece-player captured a “prey” piece-player, the predator’s hunger level reset to 0, indicating that it has previously eaten 0 moves ago (i.e. the current move).

The turn characteristic represented how many moves ago a piece-player’s last move occurred. With each turn on the chess board, the turn value was incremented by one. When a piece-player moved, its turn parameter reset to 0, indicating that it has previously moved 0 moves ago (i.e. the current move).

As an agent, the primary motivator for a piece-player was to fulfill its nutritional needs by capturing another piece of the opponent, preferably of higher value. In accordance with typical chess strategies, a piece-player might also be interested in improving its position to capture other pieces. At the same time, piece-players might have the goal to prevent themselves from being captured, either by staying protected in the range of a same-side piece (minimizing chances of being captured) or moving away from the targeted path of an opposing piece. Apart from a piece's self-interest, it can be motivated to attack the opposing king and defend the same-side king, as well as capture the opposing highest-valued piece.

A piece-player's sensory horizon was limited by its radius of vision. A piece-player could observe other pieces surrounding it and differentiate which side it belonged to. Also, a piece-player could detect the type of a piece near it, as well as the latter's distance away and location. A piece-player was apprised of how each piece moves (including itself) according to the official chess rules (FIDE 2023). Using the information about "moves", a piece-player thus had information about what squares other pieces could attack or defend (within the piece-player's field of vision).

Pieces here did not have memory of previous moves and could not think multiple moves ahead. However, before making the next move, a piece-player was programmed to disclose information about the locations of opposing pieces within its field of view to other same-team pieces (long range communication) by providing the coordinates of the opposing pieces. For example, if the opposing king was within the radius of a piece-player, then the piece-player would communicate this information to the other piece-players who also opposed that king. Conversely, if the same-side king detected it was in danger (i.e. that there was an opposing piece within its radius), the king would communicate this information to the other piece-players on that king's team so they could potentially advance closer to protect the king. Every piece-player was programmed to communicate accurate information to other pieces (by relaying from one piece to another), and to receive accurate information automatically without conflicts or errors. Information was received simultaneously and in parallel, before any move was made. While the distributed pieces took their own well-being into account, they could also determine their course of action based on information transferred from other pieces. In addition, each piece knew its own value and the value of other pieces. A piece's decision to capture was therefore influenced by the comparison of the value of itself and of an opposing piece (see Figure 1 and Figure 2).

Furthermore, as each piece had its own perspective on its outside world, it was faced with the natural limitation of not being able to see the entire board, similar to the local perspectives of cells within a body and the need for biological systems to form their own perspectives, interpretations, and models of their microenvironment (Levin 2023a; Bongard & Levin 2023; Levin 2024).

2.2. Life in Chessworld: The Algorithm of a Single Game

In biological scenarios, multiple agents can effectively act at the same time. In a standard chess game, a centralized algorithm decides which of their pieces will be moved at a given turn. In our case, the distributed agents would all try to act in each time step, due to their self-interests. We implemented our simulation on a linear computer architecture by providing a turn-based scheme that regulated the order of operations for the pieces as follows. Each piece-player initially had a radius of vision R_2. It first identified whether any pieces could be captured, and if there were, the piece-player with the highest sum of hunger level and turn would capture, and if there was a tie, it would be broken in favor of the piece-player in position to capture the highest valued piece. If a piece-player could capture multiple pieces, it would capture the higher valued one. If no pieces could be captured, then the following predefined steps occurred, varying in accordance with the experiment being performed (described in subsections below): the scheme selected the next piece-player to be moved based on necessity (a rough analogy to how cognitive systems use attention and prioritization drives to choose among actions [Bongard & Levin 2023]):

1. If a piece-player was in danger (i.e., risk of being captured) and not defended, then that piece-player was prioritized first to move.

a. Among the piece-players that were in danger, the scheme prioritized piece-players who had the highest hunger value and turn.

b. Among the remaining pieces that were not in danger, the scheme prioritized piece-players who had the highest hunger value and turn. This step is done as backup for step 2 in case the pieces in danger have no safe square to move to.

2. If a piece-player's move would place it in danger, that specific move was prioritized last.

3. Once the most prioritized move had been executed,

and the opposing team responded, this process would repeat until a checkmate or stalemate occurred.

We described how these rules were applied in detail in our different experiments in the following sections. While many possible methods could be explored, we chose this logic flow as an analog to "motivation" (i.e., it models the idea that agents with the most opportunity or the most threat would be the most active and most likely to act quicker than others).

2.3. Evolution in Chessworld: How Genomes Change over Time

In the evolutionary algorithm, 13 different genes were encoded into one chromosome, described in Tables 1 and 2. The chromosome contained five positional genes that controlled a piece-player's radius of vision and how a piece-player's state (e.g. its position, hunger level, turn) would influence its next move. Additionally, each chromosome included eight action genes that control the favorability of each potential move for the piece-player. The complete move selection process is described in Figure 2. There was no developmental process here (genotype directly encodes phenotype with no generative layer or complex mapping between them).

Once the game started and the chess engine made its move, each piece-player analyzed all the information within its radius of vision. A piece-player examined its current state by considering factors such as its position, hunger level, and turn. Before a piece made a turn, it was assigned an initial move score of zero. Then, the score was updated in accordance with positional genes based on information of itself and its neighbors. The positional genes and their range of values are summarized in Table 1 [Appended at end].

To calculate the score of each piece-player's next possible moves, we took the current positional score of each piece-player and assigned it to all its possible moves, giving all possible moves a value. In addition, every action gene updated a prospective move's score based on information of the move's outcome. The action gene and their range of values are summarized in Table 2 [Appended at end]. For example, if a rook had a positional value of 50, all its possible moves get a value of 50. If a possible move would place the piece-player into danger, the move's score updated, according to Positional Gene #4 in Table 2.

The piece with the highest move score executed their move. The chess engine responded, and the cycle repeated until the game ended in checkmate or stalemate.

2.4. Implementation Details

All simulations for both portions were built using Python 3.12.0. The opposing side to the distributed piece-players was controlled by Stockfish 15.1, winx64 avx2 version (The Stockfish developers, n.d.). To integrate the Stockfish engine with Python, the Python library (Zhelyabuzhsky 2022) was used. To ensure the engine selects moves quickly and accurately, the Stockfish engine was set to a depth of 8, and a skill level of 20. The Stockfish engine selected the third best move (average ELO [Elo 2008] of 300) for the manual portion, and randomly selected either the second best move or the third best move for the evolution portion to increase difficulty (average ELO of 500). This arrangement matched the skill level of the piece-players to ensure no side overpowers the other. We used a standard 8x8 chess board with standard chess rules, except for the En Passant (Chess.com 2024b) rule for simplicity. One game took approximately 0.8 seconds to compute. One trial consisted of 50 chess matches between the Stockfish chess engine and the distributed pieces and lasted around 40 seconds to be completed. One generation for the evolution component took on average 3 minutes. For the manual experiments, 10 trials were conducted for every experiment setting, taking 40 seconds for each. Trial results were recorded, and the mean along with standard deviation was taken for each experiment.

In the evolutionary algorithm, the skill level of the engine gradually increased every 25 generations by 1 from level 0 to level 20. The winning percentage of the piece-players for each trial served as the fitness value and was expected to approach a fitness value of 1 (100%). PyGAD 3.2.0 was used for the genetic evolution (Gad 2023). We used a population size of 200 trials per 500 generations and calculated the fitness of each trial in parallel. The genetic evolution was conducted on a Linux server, running on 2x AMD EPYC 7532, 32x2 core, 64x2 thread, 512GB RAM. In the genetic algorithm, the random mutation probability parameter was set to 10% and the single-point crossover parameter was enabled. Piece-players in each trial were characterized with a chromosome of 13 genes, controlling the influence of each policy in its decision-making.

3. Results

3.1. Minimal Distributed Agents Can Play Chess

We first established a baseline in which pieces guided by totally random legal actions played against the Stockfish chess engine for 50 matches. As might be expected, the random moving pieces had a 0% winning percentage against the Stockfish engine out of 50 chess matches, playing at an Elo of around 0 (Chess.com 2024b)—roughly the level of a person who only knows how pieces move and knows no strategy, but all moves were legal.

To determine whether distributed agents could play chess against a classical (unified) player, we then had the distributed chess pieces play against the Stockfish engine using the logic described in the manual policies portion of the Methods (see subsection “Life in Chessworld: The Algorithm of a Single Game”) and a radius of R_2. The distributed pieces were able to obtain an average win rate of 20%±6% in the manual experiments alone. After four evolutions running in parallel, the pieces were able to achieve a peak winning percentage of 44%±2%. The distributed pieces played at a maximum Elo of 1050 and an average Elo of 750 (Chess.com 2024b), equivalent to a casual chess player who knows the rules and is familiar with basic strategy. The pieces were successfully able to form defensive structures and avoid active dangers (Figure 3A). However, the pieces performed poorly near the end of the game when encircling and checkmating the king (Figure 3B). Often, the large quantity of pieces eliminated any paths for the opposing king to go, resulting in a stalemate. In addition, an inadvertent blunder of a significant piece like the queen (because of the limited radius of vision, or poor risk taking) at the beginning of the matches would make the distributed pieces’ chances of winning significantly worse. Computationally, the Stockfish engine at a depth of 8 determined the next move in around 5.8 milliseconds, while the distributed piece-players collectively decided on their next move in a total of 2.0 milliseconds.

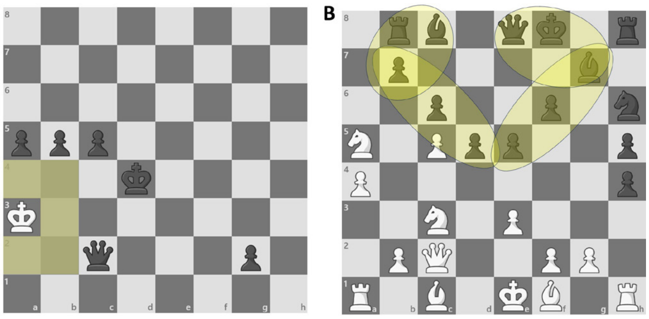


Figure 3: Scenarios of distributed agents in action. A) The black pieces, played by the distributed pieces, have a significant material advantage over the white pieces. However, the distributed pieces inadvertently force a stalemate by eliminating any squares the white king can move to. B) The black pieces, played by the distributed pieces, create a defensive structure, forming two pawn chains of length three. A black rook and black bishop defended a pawn at the b7 square, a black bishop defends the black pawn on the f6 square. The black queen protects three pawns, and a bishop, creating a defensive front.

3.2. Radius of Vision Positively Correlates with Winning Percentage

We next sought to understand whether, and to what degree, being able to see further along the board would enable the collective to play better. Thus, to analyze the correlation of radius of vision and winning percentage in the manual experiments, we altered the radius settings (8, between R_0 and R_7) to analyze which was most optimal.

The mean and standard deviation of the data are shown in Figure 4A [Appended at end]. We found that the winning percentage sharply increased from radii R_0 to R_2, but plateaued for radii R_2–R_7 with minor fluctuations. The highest winning percentage became apparent from a radius of R_4, with a 21%±3%. We conclude that a larger radius of vision improves performance for all pieces, but the gains drop off and having information about distant regions of the board does not add much to the efficacy of play under these conditions.

We then tested the evolutionary approach, setting possible radius bounds for alleles in the population to range from R_2 to R_7 inclusively. Initially, the frequency of each radius was randomized (Figure 4B). Throughout the evolution, the radius of the best trial in each generation fluctuated between R_3–R_7. At the end of the evolution, R_4 appeared the most frequently in the final population (as seen in Figure 4C), with 27%±15% of trials in the final population having an R_4 radius. The evolution had a higher winning percentage (44%±2%) than the manual experiments’ results because the evolution balanced all the policies together, complementing the optimal radius (R_4). Comparing Figure 4B and Figure 4C, there is a contrast between the frequency of radii at the beginning of the evolution, and at the end [Appended at end].

3.3. Long Range Information Transfer Positively Correlates with Winning Percentage

To determine the benefits of information transfer between piece-players, a long-range communication rule was implemented in the manual experiments during a turn only when the opposing king was within the radius of a piece-player. We utilized the base scheme described in Methods section. Before moves were sorted based on whether a move would place a piece-player in danger, moves that increased the distance between the piece and the opposing king were prioritized last. If the opposing king was not in the radius of a piece, the pieces would play according to the base scheme. Keeping the radius to R_2, we compared the winning percentages of the winning performance without the new rule and with the new rule. We observed that R_2 had a winning percentage of 20%±4%. The winning percentage of the experiment with the new rule increased to 24%±6% (Figure 5A) [Appended at end]. (1-tailed, two-sample equal variance [homoscedastic] T-test, p=0.066). Thus, we observed that the long-range communication group's winning percentage was 4% more than the control group.

We then tested the evolutionary approach with the same goal, by comparing the frequency of genes in the initial and final population. Genes related to long-range information transfer had parameter values between -100-100 (see Policy 6, 7 and 8 in Table 2). The parameter values of each trial in a generation were grouped in ten bins, each bin of size 20 (as seen in Figure 5B-5G) to capture the main features of the data. The results of the genetic algorithm revealed that genes related to long-range information transfer were reliably selected for in the population with the best fitness.

Starting with the gene that controlled whether a piece-player prioritized moving closer to the opposing king (see Policy 6 in Table 2), we observed that on average, 29%±14% of parameter values in the final population were concentrated between 20 and 40, while 18%±13% of values were concentrated between 0 and 20 in the four evolutions (Figure 5C), deviating from the random trend in the initial population (Figure 5B). The most successful individuals in each of the four runs had parameter values of 7, 21, 23, and 40, showing the prioritizing moving closer to the opposing king (Policy 6 in Table 2) was favorable in the decision-making of a piece.

Then, we analyzed the gene that controls whether a piece-player prioritized moving closer to the defending king if an opposing piece is in the defending king's radius of vision (see Policy 7 in Table 2). We observed that on average, 32%±15% of parameter values in the final population were concentrated between -40 and -20 (Figure 5E). The most successful individuals in each of the four runs had parameter values of -8, -9, -29, and -38, deviating from the random trend in the initial population (Figure 5D). This implies that prioritizing moving closer to the same-side king (Policy 7 in Table 2) was unfavorable in the decision-making of a piece.

Then, we analyzed the gene that controls whether a piece-player prioritized moving closer to the highest-valued piece (see Policy 8 in Table 2). We observed that on average, 22%±11% of parameter values in the final population were concentrated between 20 and 40 (Figure 5G), deviating from the random trend in the initial population (Figure 5F). The most successful individual within the population in the four runs had parameter values of -5, 34, 49, and 96, demonstrating that prioritizing moving closer to the highest-valued piece (Policy 8 in Table 2) was favorable in the decision-making of a piece.

From the results, the ability to transfer long-range information about the opposing king and the highest valued piece proved to be highly favorable. Defending the same-side king proved to be unfavorable, as the parameter values were largely negative, also showing that the pieces fared better on the offensive.

3.4. Courage is More Favorable than Caution in Certain Scenarios

We next sought to understand the contribution of risk-taking and risk aversion to the quality of play by the collective agent. To determine whether pieces should act more courageously or cautiously for the best winning percentage, we analyzed the performance difference between these two strategies. A piece was cautious when it deliberately avoided or prevented the risk of being captured by choosing an alternative safe move. A piece was courageous when exposing itself to risk.

To implement these strategies in the manual experiments, we utilized the base scheme described in Methods section. However, we changed how pieces made decisions based on danger in two experiments. The first experiment examined when pieces acted cautiously, following the base scheme. The second experiment showcased when a piece acted courageously for accepting the risk of being captured. This experiment is the same as the first one, except a piece-player is not considered to be in danger if it was defended by another same-side piece-player. In addition, a piece-player's move is not considered dangerous if a piece would be (after the potential move) defended by another same-side piece. We observed (Figure 6A) [Appended at end]. that the winning percentage was 18%±6% when the piece had only a sense of danger. When the piece could detect that it was defended and in danger at the same time (resulting in no action), the winning percentage was 15%±5%. We see that the more cautious pieces have a mildly higher winning percentage than the courageous ones.

To have a more accurate understanding of which specific courageous and cautious strategies are optimal, we allowed evolution to set the risk-taking level for the pieces. Genes related to courage and caution had parameter values between -100-100 (see Policy 2, 5 in Table 1 and Policy 4, 5 in Table 2). The parameter values of each trial in a generation were grouped in ten bins, each bin of size 20 (as seen in Figure 6B-6I) to capture the main features of the data. The results of the genetic algorithm revealed that genes that encouraged more risk were more favorable. Starting with the gene that controls how motivated a piece is to escape imminent danger (see Policy 2 in Table 1), we observed that 62%±14% of parameter values in the final population were between 20 and 40, (Figure 6C), deviating from the random trend in the initial population (Figure 6B). The most successful individual within the population in the four runs had parameter values of 12, 12, 20, 22, demonstrating that the gene had a large influence on the decision-making of a piece.

We then analyzed the gene influencing the decision-making for a piece when the piece is defended by another same-side piece (see Policy 5 in Table 1). 76%±4% of parameter values in the final population were concentrated around 0-20 (Figure 6E), deviating from the random trend in the initial population (Figure 6D). The most successful individual within the population in the four runs had a weighting of 3, 7, 9, and 18, indicating that the gene had some influence in the decision-making.

We then analyzed the gene that controls voluntary decision of a piece to put itself in danger (see Policy 4 in Table 2). We observed that 83%±1% of the parameter values in the final population on average were between 0-40 among four evolutions (Figure 6G), deviating from the random trend in the initial population (Figure 6F). The most successful individual within the population in the four runs had a weighting of 1, 2, 2, and 3, highlighting that this policy had some influence in the decision-making of a piece.

We then analyzed the gene influencing the decision-making for a piece when the piece's next move would be defended by another same-side piece (see Policy 5 in Table 2). 81%±5% of the parameter values in the final population were between 0-20 among four evolutions (Figure 6I), deviating from the random trend in the initial population (Figure 6H). The most successful individual within the population in the four runs had a weighting of 5, 23, 24, and 34, highlighting that this policy had some influence in the decision-making of a piece, and making it more cautious.

The results show that piece-players that were at risk of being captured were more eager to move than piece-players that were not at risk of being captured, meaning piece-players were cautious. In addition, piece-players in general acted courageously and moved to squares that could lead them to be captured (Figure 6J). Genes depending on whether a piece-player is protected and whether a piece's move would be protected by other pieces had minor influence on the decision-making about the next move, however still favoring protection over risk. Thus, piece-players were cautious about their position when they were at risk of capture, and were courageous when making a move.

3.5. Having More Patience when Hungry, and Less Patience when Not Having Moved in a While, is Optimal

The ability to perform "delayed gratification" in a problem space-making moves that temporarily take the agent further from its goals in order to recoup gains later—is one metric of basal intelligence (James 1890). It is interesting to ask what kind of policy should be used among the components of a collective intelligence to determine which ones get to act at what time, for optimal adaptive performance. Thus, we next sought to determine what was the optimal move order for the pieces, we analyzed which strategy is more optimal—making decisions based on the hunger level, turn, or both. We compared the results to when neither of the attributes are applied. A piece-player was considered to be more patient when its hunger level and/or turn was high, but the piece refrained from moving. A piece was considered to be less patient if its hunger level and/or turn was high, and the piece prioritized its self-interests and moved.

For our first experiment, we prioritized pieces with the highest hunger value and turn (as described in the base scheme). For the second experiment, we prioritized neither of these values instead. For the third experiment, we prioritized pieces with the highest hunger value, and for the fourth experiment, we prioritized pieces with the highest turn instead. Keeping the radius to R_2, we compared the winning percentages of the winning performance of each of the moves' ordering methods. From the four experiments, we observed that the control had a 14%±5% winning percentage, the hunger based moving pieces had a 10%±3% winning percentage, and the alternating order moving pieces had a 16%±4% winning percentage (Figure 7A) [Appended at end]. It appears the collective did best when emphasizing a strict turn order for its members.

To analyze which strategies were most optimal and their magnitude of impact, we conducted the evolutionary by comparing the frequency of genes in the initial and final population. Genes related to hunger and turn had parameter values between 0.0-5.0 (see Policies 3 and 4 in Table 1). The parameter values of each trial in a generation were grouped in ten bins, each bin of size 0.5 (as seen in Figure 7B-7E) to capture the main features of the data. The results of the genetic algorithm revealed that having more patience when hungry and less patience when not having moved in a while is optimal.

For the gene controlling how motivated a piece is based on their hunger (see Policy 3 in Table 1), we observed that on average, 77%±6% of the parameter values in the final population were concentrated around 0-0.5 (Figure 7C), deviating from the random trend in the initial population (Figure 7B). The most successful individual within the population in the four runs had parameter values of 0.05, 0.1, 0.1, and 0.2, demonstrating that a piece ignoring their hunger completely is optimal.

Then, we analyzed the gene that controls how motivated a piece is based on their turn (see Policy 4 in Table 1). We observed on average, 36%±14% of the parameter values in the final population were concentrated around 3.50-4.00, and ~60% of values were above 3.5, with the maximum possible parameter value being 5 (Figure 7E), deviating from the random trend in the initial population (Figure 7D). The most successful individual within the population in the four runs had parameter values of 2.7, 3.4, 3.6, and 3.6, indicating that the policy had a large impact on the decision-making of a piece.

The hunger level multiplier clearly indicates that the hunger level was almost negligible in the decision-making process, meaning it is best when piece-players are patient with their hunger when making decisions. Conversely, the turn multiplier shows that prioritizing to make a turn after not doing so in a while is deemed to be more beneficial.

3.6. Adding a "Threatening" Drive Significantly Improves Performance

In the basic scheme, the only drive that guides pieces' behavior is the ability to consume another piece. We next sought to examine the consequences of giving them a motivation to threaten another piece. Moves were prioritized based on whether they would place an opposing piece in danger. We varied the radius of vision from R_0 to R_7 and compared the winning percentage to the original base scheme (Fig. 4A). The winning percentages of this experiment were significantly higher than the base scheme (except for R_0 and R_1), with higher radius of vision performing better. For example, the winning percentage for R_7 with the new rule was 42%±5%, while the winning percentage for R_7 without the new rule was 20%±6% (1-tailed, two-sample equal variance (homoscedastic) T-test, $p<<0.01$). The ELO improved by ~50.

4. Discussion

We tested the hypothesis that a passable game of chess could be played without a central planner, memory, training in prior games, forethought, or consideration of the consequences of specific actions. By implementing a bottom-up, distributed player where the pieces had their own agency, we created an alternative to conventional chess AI (Duca Iliescu 2020; K.B. 2021)—one based on the concepts of collective intelligence (Couzin 2007; Couzin 2009; Couzin *et al.* 2002; Deisboeck & Cousin 2009; McMillen & Levin 2024; Witkowski & Ikegami 2019; Pinero & Sole 2019; Sole *et al.* 2016; Heylighen 2013; Wheeler 1911; Ward *et al.* 2008; Bazazi *et al.* 2008). We found that this very minimal system operates at the level of a human novice when each piece is computing its own desired moves. Using evolutionary algorithms with only 13 possible parameters, a ragtag group of player-pieces can in aggregate increase its score to an ELO of 1050, which is comparable to the score needed to be competitive against a beginner.

4.1. Parameters Impacting Play Quality

A critical component of any collective intelligence is the set of policies which regulate their actions and interactions. As with many examples of collective problem-solving, the parameter values that improve the functionality of the group are not obvious and hard to predict from first principles (McMillen & Levin 2024; Rahwan 2019). Our exploration of the parameter space and evolutionary search found several ways to optimize the performance; First, that the optimal radius of vision was R4. The reason why it was not significantly higher (like R6 or R7) was most likely because a too large radius of vision can hamper a piece's decision-making: with too much awareness of opposing pieces across the board, a piece may move to aggressively, stranded in enemy territory and ultimately sacrificed. Conversely, a lower radius of vision that 4 performed significantly worse, presumably because unaware of the whereabouts of opponent's pieces, therefore diminishing captures and diminishing the winning percentage of the collective. The radius of R4 allows for the balance between controlled risk taking and capture.

We found that the best performance came when pieces were more cautious about their current position, i.e. escape imminent danger if an opposing piece is attacking them, but more courageous (i.e. take more risks) when selecting their next move. They avoided imminent threats, regardless of whether they were defended by a same-side piece. The pieces were content with placing themselves at risk while moving to squares where they would be protected, thus expanding the position and working as a collective. This configuration allows offensive attacks, while preventing passive play and takes into consideration the present danger/ defense set up.

For long-range communication, the policy of having knowledge of the opposing king's position and the opposing highest valued piece allowed for swarming. This ability was influential, which is reflective by the increase in winning percentage in the manual experiments. The ability to defend the same-side king had a negative influence in the decision-making process. This is because defending the same-side king promotes passive play, allowing the chess engine to take down the distributed agents. As a result of these policies, pieces were able to collectively surround the opposing king to deliver checkmates more consistently, and attack the opposing highest valued piece, all while playing offensively.

Moves driven by hunger level were seen to be suboptimal. In the manually-parametrized experiments, moving based on hunger level decreased the winning percentage compared to the other moves' ordering strategies. Moreover, the hunger level multiplier in the evolution portion was almost at its minimum, because hunger level provided little input in the decision-making for the next move. On the other hand, having the desire to move when having not moved in a while (based on turn) was favored significantly more in the evolution, and moving based off turn boosted the winning percentage. In classical chess theory, it is typically suggested to move a wide range of pieces to develop and strengthen the chess position, which is what the distributed agents confirmed.

4.2. Emergent Collective Goals

One formalism for the study of collective intelligence is the notion of the cognitive light cone—the spatio-temporal radius of the largest state that an agent can actively work towards (Levin 2019). Here, pieces are limited by their radius of vision. However, when pieces are able to communicate with one another, they are able to expand their spatial cognition and receive information about pieces across the board when applicable. An example of this is knowing the location of the opposing king. If a piece were to attempt to target an opposing king in their radius of vision, the probability of the king stepping into the radius of vision (e.g. R_4) is not high, making the policy ineffective, with few pieces being able to know the location of the opposing king at a given time. When pieces are able to communicate to other pieces about the location of the opposing king, pieces are able to create consistent pressure throughout the match and checkmate the opposing king. This expansion of cognition is significant, in that a piece does not require its own large radius of vision like R_7 to successfully apply pressure. A smaller radius of vision is sufficient for a piece's needs, because the collective communication provides a larger effective sensory radius—like that seen in group sensing in weakly-electric fish who can effectively "see" through each other's senses (Pedraja & Sawtell 2024).

Each piece-player has its immediate goal to survive by working to capture valuable opposing pieces. From pawn to queen, every piece player moves and positions themselves in such a way that satisfies its metabolic instincts. However, transcending the level of the individual, the functional purpose of the collective is ultimately to checkmate the opposing king. This is an emergent outcome, not specifically encoded in the algorithm. Despite having desires and constraints that occasionally hamper play, the individuals' primary motivations (e.g. capturing opposing pieces) align with the collective's overarching goal. These motivations converge into one effort, thereby boosting the resiliency of the collective and allowing them to reach their goal in the face of internal adversity.

One fascinating question, bearing on discussions of whether intelligence is intrinsic or observer-dependent (Bongard & Levin 2023), is: what does an external observer, who knows nothing about the inner construction of each player, think of the games that our swarm plays? Would a chess-savvy observer see game-level goals being pursued—emergent long-term strategies in the eye of the beholder that do not exist in the ground truth of the algorithms being pursued by the agents (Heider & Simmel 1944; Scholl & Tremoulet 2000)? We recently showed a similar phenomenon in sorting algorithms (Zhang, Goldstein, & Levin 2024), which were exhibiting several behavioral problem-solving traits that had not been baked in to their algorithm directly (Zhang, Goldstein, & Levin 2024). In our dataset, what could be observed were: pawns marching forward despite not having the knowledge of queen promotion, especially the center pawns, the queen and pieces in the center of the board were active at the very beginning of a chess game, and pieces went on the attack (offensive), venturing to the opposite side of the board. They did not play passively.

Central to the function of collectives are the balances of cooperation and competition among their members (Gawne, McKenna, & Levin 2020; Strassmann & Queller 2010). The distributed pieces in the Chessworld might have two types of conflict, intra-pieces conflict, and inter-pieces conflict. Intra-piece conflict encapsulates the conflict of self-interests between the pieces. More notably, factors like nutritional needs, patience, and protection (all ingrained in each piece) may cause conflict among pieces (only emergent in gameplay) in the decision to move or avoid moving. Inter-pieces conflict expresses the individual piece's disunity with the group's goals. A piece's desire might not be in accordance with the team's goals. For example, a piece might be used as a sacrifice, or be prevented from moving due to the strategic position on the board. These inter-pieces conflicts appear in gameplay depending on the game dynamics.

4.3. How Does Bottom-up Chess Play Compare to Human Players? An Informal Analysis

One of the authors (GC) is a former chess player (max. Elo rating = 2270) with 15 years of chess coaching experience from absolute beginners to international masters, and a prolific researcher in chess expertise (Bilalić, McLeod, & Gobet 2007; Campitelli & Gobet 2008; Campitelli, Gobet, & Bilalić 2014). GC has played several games against the distributed piece player and observed its behavior. He made the following observations. The distributed piece engine plays like a clever 6-year-old child who has just learned the rules of the game. The engine is excellent at detecting when it can capture an opponent's piece, which is common in novice players with some experience in chess playing, but not so common in children who have just learned the rules of the game. Individual differences are typically observed, with more intelligent children detecting they can capture pieces faster than other children (Campitelli *et al.* 2007). Another characteristic of the distributed piece engine is that of following the concept of development. In chess, is a strategic concept by which a player moves several pieces at the beginning of the game; rather than moving the same piece several times (Capablanca 1921). Development is one of the first strategic concepts taught to novices (Rozman 2023), with intelligent children learning this concept faster than other children. Another characteristic of the engine that resembles. The development applied by the distributed chess engine is not optimal (i.e., it does not move the pieces to the best positions) but, again, it reflects a smart kid who, instead of moving the same piece several times discovers that it is better to develop several pieces.

An important characteristic of the distributed chess engine is its difficulty to check mate the opponent. It is capable of capturing opponent pieces, but its behavior denotes it does not have the concept of check, let alone that of check mate. Therefore, it either stalemates the opponent or check mates the opponent by chance. Again, this is a very common occurrence in children's games in which one of the players is much better than the opponent and captures all their pieces (except the King) but they struggle to checkmate the opponent.

The version of the distributed engine that contains a "threatening drive", which improved performance relative to the original engine as shown in Figure 8 [Appended at end], shows a very different "human style". The "threatening drive" version does not do piece development well because it is very keen on attacking, moving a piece to attack an opponent's piece and in the next move it moves the same piece again to capture the opponent's piece. This version does not look clever any longer as it sometimes moves the Queen to attack a pawn and captures the pawn, allowing the opponent to capture the queen. Rather, it is a very aggressive player. On the other hand, this version is a much better player at endgames in which all the pieces of the opponent are captured and the engine has to checkmate the opponent. The "threatening drive" version behaves as a child who has just learned how to check and checks the opponent all the time. Given that it checks the opponent, this version is more likely to check mate the opponent that the version without the threatening drive. Summing up, the previous version is a better and more conservative player in the opening stage, and the current version is an aggressive player during the whole game, and better player in the endgame, given that it is more likely to check mate the opponent.

4.4. Limitations of the Study

There are several aspects of the current system and dataset which will be developed and explored in subsequent work. These include additional analysis of the games to uncover novel emergent features of strategy, allowing more individual identity to the different types of pieces (specialization), and a deeper investigation of the role of scheduling in this process. In current digital architectures, it is very difficult to truly implement simultaneous actions by a swarm—while possible in the (macroscopically) continuous 3D world, standard architectures must break down the moves into atomic operations, preventing truly independent activity. The role of these dynamics in the outcomes must be studied more deeply and examined in parallel architectures. Likewise, the implementation of negotiation among the pieces could enrich game-theoretic perspectives and evolutionary dynamics.

Additional future work will be focused around finding ways that improve play further while maintaining the minimal nature of the agents. For example, we recently suggested the role of stress sharing as another kind of cognitive glue (Shreesha & Levin 2024); this and other biological dynamics will be explored. Finally, it will be important to extend this approach to other classic games (checkers, Go, etc.) beyond our analysis of chess, to see where it is successful and what game conditions are or are not ideal for a distributed approach.

Conclusions

It is tempting to draw categorical distinctions between swarms and "true unified beings" like human beings and other brainy organisms. However, all of us are made of parts and all intelligences are, in a sense, collective intelligences. Even human beings are made of components which must work together to result in a degree of unified performance (Sole, Moses, & Forrest 2019; Pinero & Sole 2019; Seoane 2019; Martinez-Corral *et al.* 2019; Manicka & Levin 2019)—collective dynamics which exhibit occasional breakdowns, resulting in cognitive dissociation or morphogenetic dissociation disorder known as cancer (Braude 1995; Levin 2019; Levin 2021). Thus, there are no truly unified, monolithic, monadic chess players, and our individual neurons likely do not know about the strategies of chess any more than our in silico virtual players do. However, neurobiological studies of novices and chess grandmasters have revealed differences—specifically, increased whole-brain functional connectivity patterns (Song *et al.* 2022; Liang *et al.* 2022; Amidzic *et al.* 2001). Thus, using information theory to understand the relationship between parts and whole (Kolchinsky *et al.* 2014; Sporns 2011; Bullmore & Sporns 2009; Tononi, Edelman, & Sporns 1998; Tononi, Sporns, & Edelman

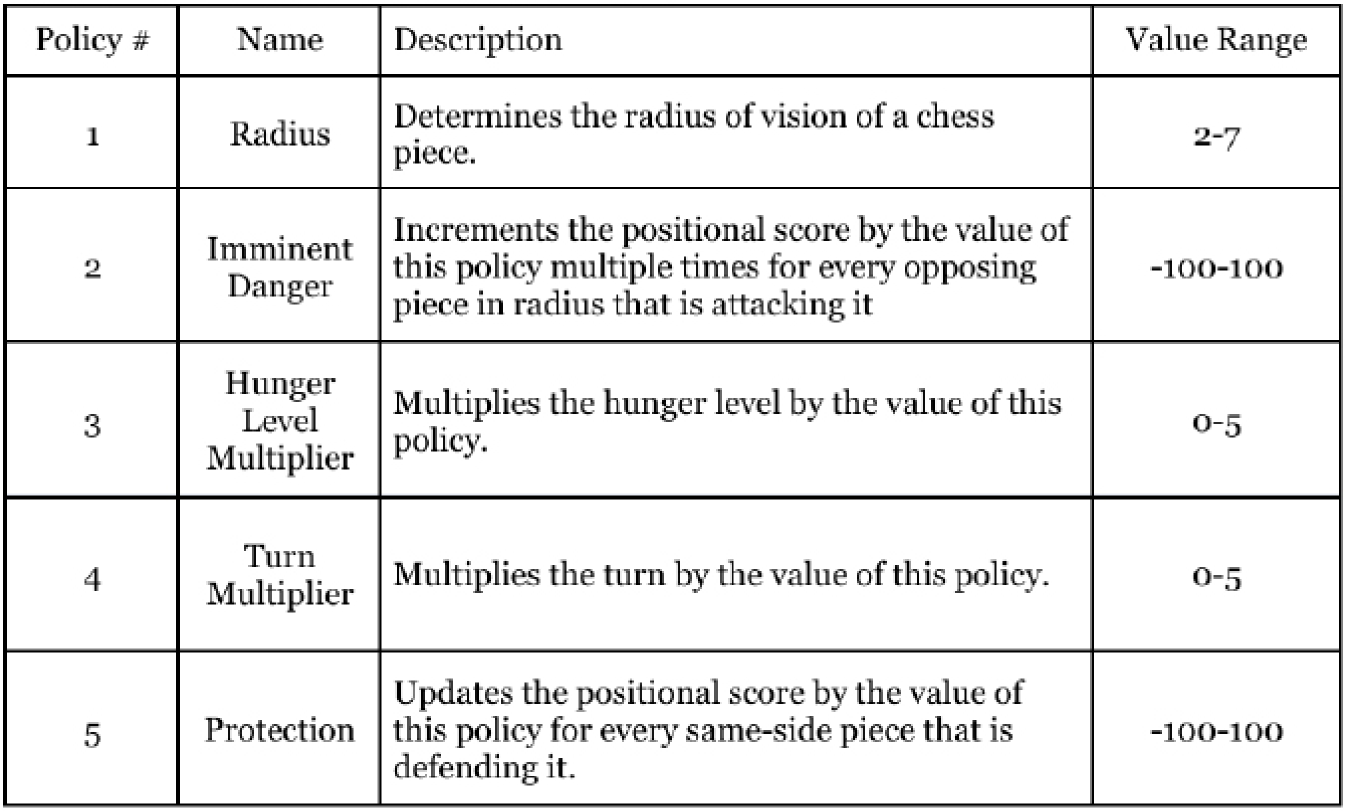
1994; Albantakis *et al.* 2017; Hoel *et al.* 2016; Hoel, Albantakis, & Tononi 2013) in minimal model systems and strongly constrained virtual worlds are likely to enable rich comparison between artificial life and natural biological beings. It is also interesting however that while in our simulation, both the individual agents and the collective intelligence both live in the same world (Chessworld), biological collective intelligences project themselves into new worlds, as evolution pivots the tools needed to navigate physiological and gene expression spaces into anatomical morphospace, 3D behavioral space, linguistic space, and many others (Fields & Levin 2022; Levin 2023).

We believe it is essential to develop a science not only of emergent complexity (Adami 2002; Prokopenko, Boschietti, & Ryan 2009), but of emergent cognition: to be able to predict the appearance of, and characterize the problem-solving competency and effective goals of, novel unconventional agents such as swarms of robots or minimal active matter (Blackiston *et al.* 2023; Strong, Holderbaum, & Hayashi 2024; Adamatzky, Chiolerio, & Szacilowski, 2020; Cejkova *et al.* 2017; Hanczyc 2014), of large-scale financial and political structures constructed in societies, and in the collective intelligence of our own brains, composed of large numbers of competent cells which nevertheless give rise to problem-solving, forward-thinking beings (Chater 2018; Seth 2013) (Tononi, Edelman, & Sporns 1998; Friston 2013; Ramstead *et al.* 2022) with many unanswered questions about our nature, our capabilities, and ways in which those supervene on the biochemistry and physiology of our components.

Acknowledgments

We thank Adam Goldstein for many helpful discussions and comments on the manuscript, James Buchanan for discussions emphasizing the importance of the surrounding harness when studying the emergent cognition in simulations of minimal agents, and Hananel Hazan for technical support with HPC Linux servers. M.L. gratefully acknowledges support via Grant 62212 from the John Templeton Foundation. We thank Julia Poirier for assistance with the manuscript.

Table 1. Positional genes and parameters



Policy #	Name	Description	Value Range
1	Radius	Determines the radius of vision of a chess piece.	2-7
2	Imminent Danger	Increments the positional score by the value of this policy multiple times for every opposing piece in radius that is attacking it	-100-100
3	Hunger Level Multiplier	Multiplies the hunger level by the value of this policy.	0-5
4	Turn Multiplier	Multiplies the turn by the value of this policy.	0-5
5	Protection	Updates the positional score by the value of this policy for every same-side piece that is defending it.	-100-100

Table 2. Action genes and behavioral parameters their loci determine

Policy #	Name	Description	Value Range
1	Capture	Increments the move score by the value of this locus only if a piece can capture another piece.	-100-100
2	Favorable capture	Increments the move score by the value of this locus only if a piece can capture an opposing piece of higher or equal value.	-100-100
3	Unfavorable Capture	Updates the move score by the value of this locus only if a piece can capture an opposing piece of lower value.	-100-100
4	Dangerous Move	Updates the move score by the value of this locus if the move places a piece in danger. The score is updated multiple times for every opposing piece in radius that would be attacking it.	-100-100
5	Defended Move	Increments the move score by the value of this locus if a same-side piece will defend the moved piece. The score is updated multiple times for every same-side piece that would be defending.	-100-100
6	Approach Opposing King	If the opposing king is in the radius of any distributed piece and the move brings the piece closer to the opposing king, it increments the move score with the value of this locus.	-100-100
7	Move Closer to Same Side King	If the same-side king is in danger (an opposing piece is within its radius) and the move brings the piece closer to the same-side king, the value of this locus increments the move score.	-100-100
8	Approach Highest Valued Opposing Piece	If the move brings the piece closer to the opposing highest-valued piece, the value of this locus increments the move score.	-100-100

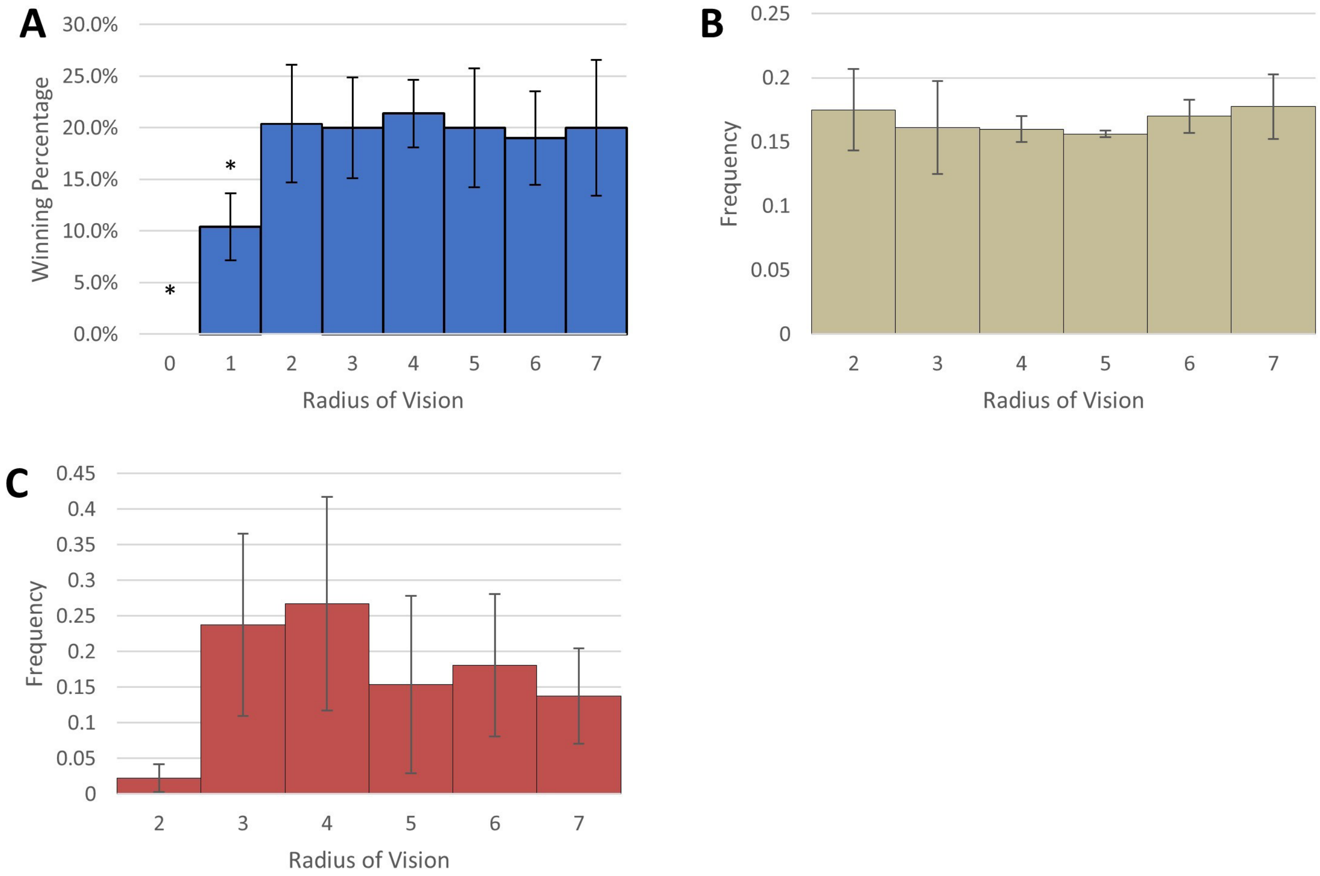
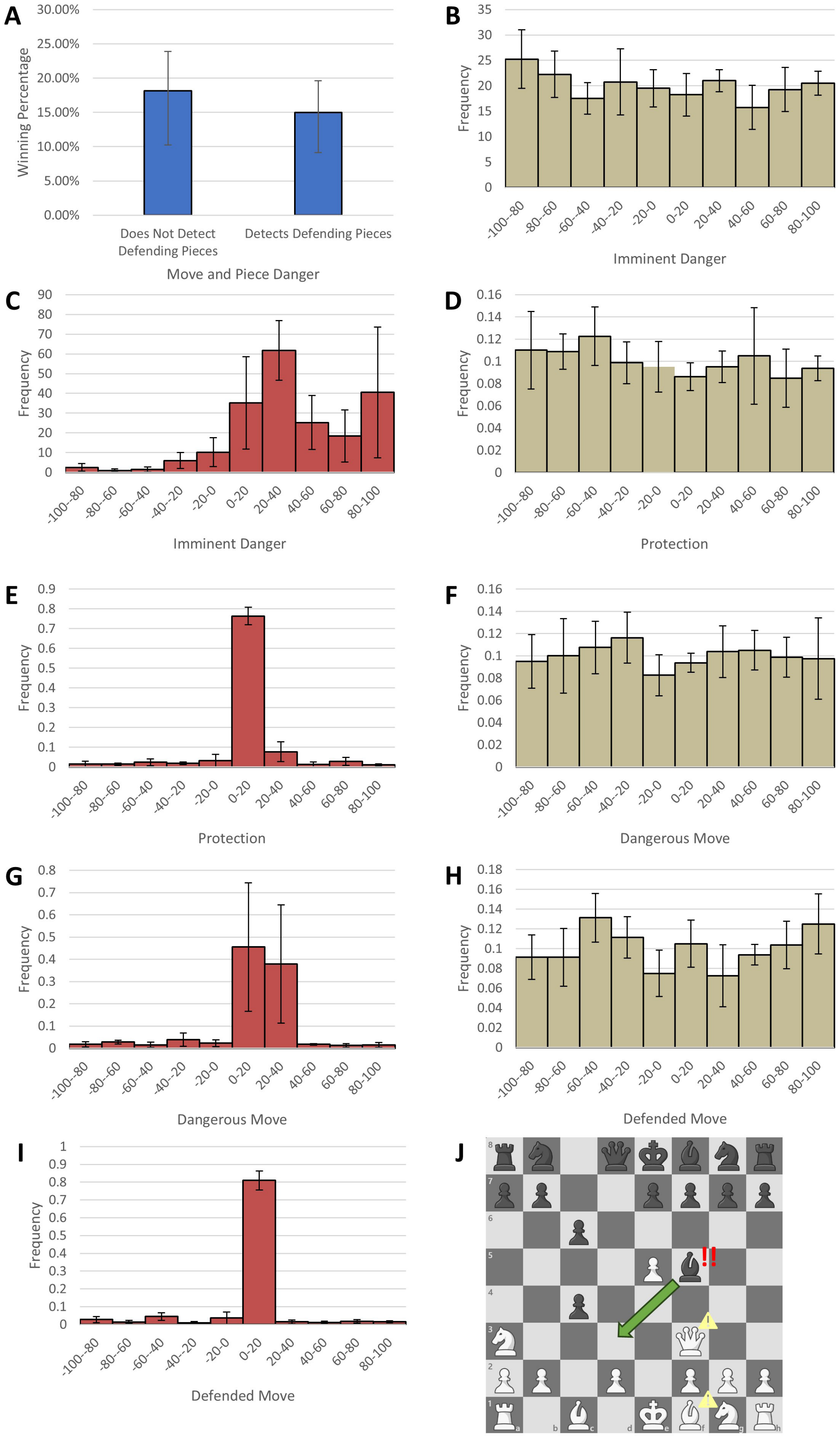


Figure 4: Effect of range of vision on play quality. (A) Winning percentage by players using each radius of vision (N=50). R_0 and R_1 winning percentages were significantly lower than R_2 (B) Initial frequency distribution of radius values (N=200) across the population before evolution for Radius of Vision (Policy 1 in Table 1) between R_2 and R_7. (C) Final frequency distribution of radius values (N=200) across the population before evolution for Radius of Vision (Policy 1 in Table 1) between R_2 and R_7.

Figure 5 (next page): Long Range Communication among pieces increases winning percentage. (A) Comparison between the control group, and the group with long range communication (N=50). The control group not including long range communication had a winning percentage of 20%±4% while the group including long range communication had a winning percentage of 24%±6%. T-test resulted in a 0.066 significance value. (B) Initial frequency distribution of parameter values (N=200) across the population before evolution for the gene that controls whether a piece moves closer to the opposing king (Policy 6 in Table 2). Parameter values were randomized, resulting in bars of similar size (C) Final frequency distribution of parameter values (N=200) across the population after the evolution for the gene that controls whether a piece moves closer to the opposing king (Policy 6 in Table 2). 29%±14% of parameter values in the final population were concentrated between 20 and 40, while 18%±13% of values were concentrated between 0 and 20. (D) Initial frequency distribution respectively of parameter values (N=200) across the population for the gene controlling whether a piece moves closer to the defending king if an opposing piece is in the defending king's radius of vision (see Policy 7 in Table 2). (E) Final frequency distribution respectively of parameter values (N=200) across the population for the gene controlling whether a piece moves closer to the defending king if an opposing piece is in the defending king's radius of vision (Policy 7 in Table 2). (F) Initial frequency of parameter values (N=200) respectively for the gene controlling whether a piece moves closer to the highest-valued piece (Policy 8 in Table 2). (G) Final frequency of parameter values (N=200) respectively for the gene controlling whether a piece moves closer to the highest-valued piece (see Policy 8 in Table 2).


A
Winning Percentage
30.00%
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
Does Not Detect Defending Pieces
Detects Defending Pieces
Move and Piece Danger
B
Frequency
Imminent Danger
C
Frequency
Imminent Danger
D
Frequency
Protection
E
Frequency
Protection
F
Frequency
Dangerous Move
G
Frequency
Dangerous Move
H
Frequency
Defended Move
I
Frequency
Defended Move
-100--80
-80--60
-60--40
-40--20
-20-0
0-20
20-40
40-60
60-80
80-100
J
!!
a b c d e f g h
1 2 3 4 5 6 7 8

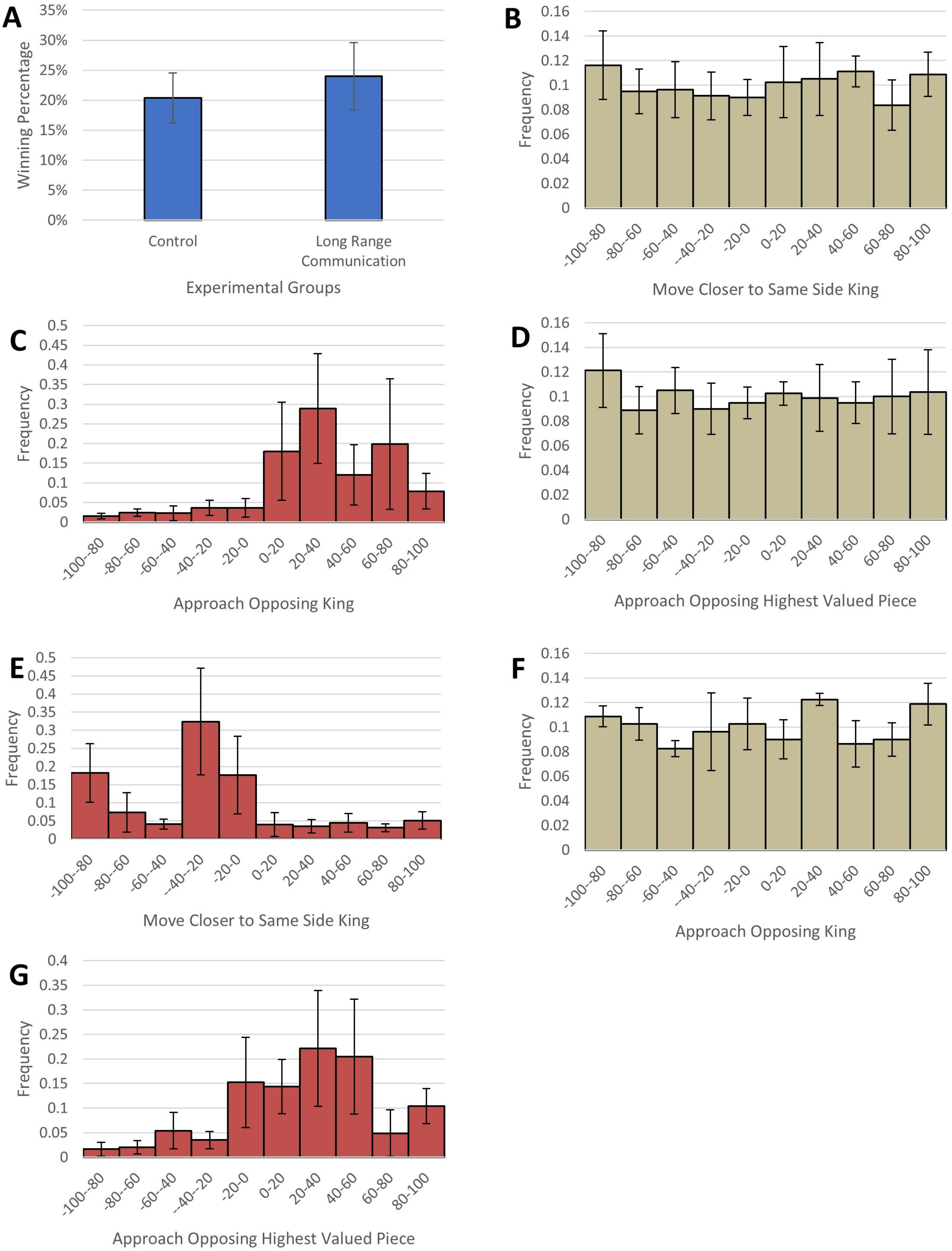


Figure 6: Courage is more favorable than caution in certain scenarios . (A) Pieces that are more cautious (denoted as ‘Does Not Detect Defending Pieces’) have a higher winning percentage than those that have more courage (denoted as ‘Detects Defending Pieces’) (N=50). (B) Initial frequency distribution of parameter values (N=200) across the population before evolution for the gene that controls how motivated a piece is to escape imminent danger (see Policy 2 in Table 1). (C) Final frequency distribution of parameter values (N=200)

across the population after the evolution for the gene that controls how motivated a piece is to escape imminent danger (see Policy 2 in Table 1). (D) Initial frequency distribution respectively of parameter values (N=200) across the population for the gene influencing the decision-making for a piece when the piece is defended by another same-side piece (see Policy 5 in Table 1). (E) Final frequency distribution respectively of parameter values (N=200) across the population for the gene influencing the decision-making for a piece when the piece is defended by another same-side piece (see Policy 5 in Table 1). (F) Initial frequency of parameter values (N=200) respectively for the gene controlling the voluntary decision of a piece to put itself in danger (see Policy 4 in Table 2). (G) Final frequency of parameter values (N=200) respectively for the gene controlling the voluntary decision of a piece to put itself in danger (see Policy 4 in Table 2). (H) Initial frequency of parameter values (N=200) respectively for the gene influencing the decision-making for a piece when the piece's next move would be defended by another same-side piece (see Policy 5 in Table 2). (I) Final frequency of parameter values (N=200) respectively for the gene influencing the decision-making for a piece when the piece's next move would be defended by another same-side piece (see Policy 5 in Table 2). (J) The black pieces are played by the distributed pieces. In this position, the black light squared bishop is under attack by the white queen and is inclined to move out of danger. The black bishop accepts the risk of moving to the square at the end of the green arrow because a same-side piece-player (black pawn) is defending that square, despite the white queen and white bishop guarding that square.

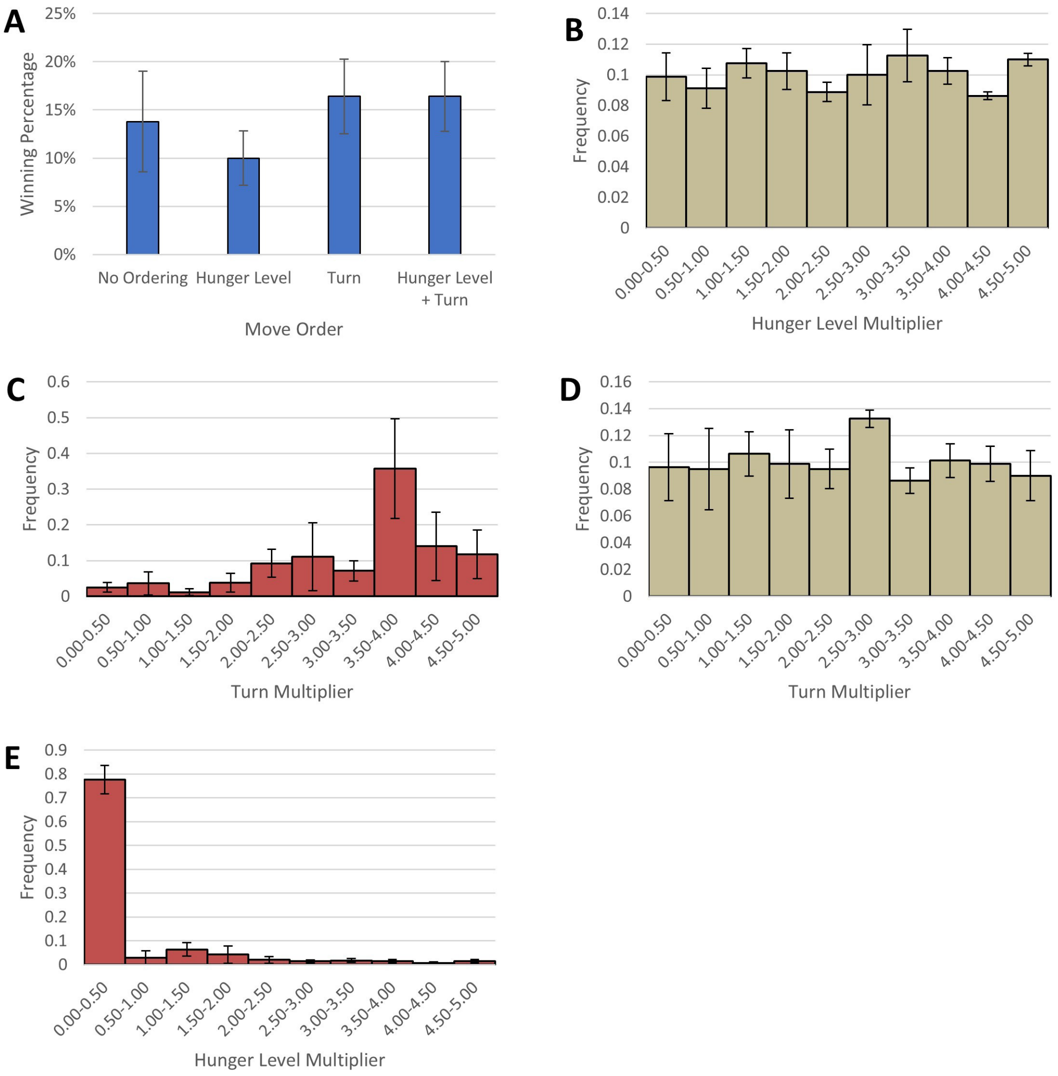


Figure 7: Quality of performance is best optimized when using strict turn order to determine activity of the agents. (A) Different ordering strategies of moves when no captures are available (N=50). (B) Initial frequency distribution of multiplier values (N=200) for the gene controlling how motivated a piece is based on their hunger (Policy 3 in Table 1). (C) Final frequency distribution of multiplier values (N=200) for the gene controlling how motivated a piece is based on their hunger (Policy 3 in Table 1). (D) Initial frequency distribution of multiplier values (N=200) for the gene that controls how motivated a piece is based on their turn (Policy 4 in Table 1). (E) Final frequency distribution of multiplier values (N=200) for the gene that controls how motivated a piece is based on their turn (Policy 4 in Table 1).

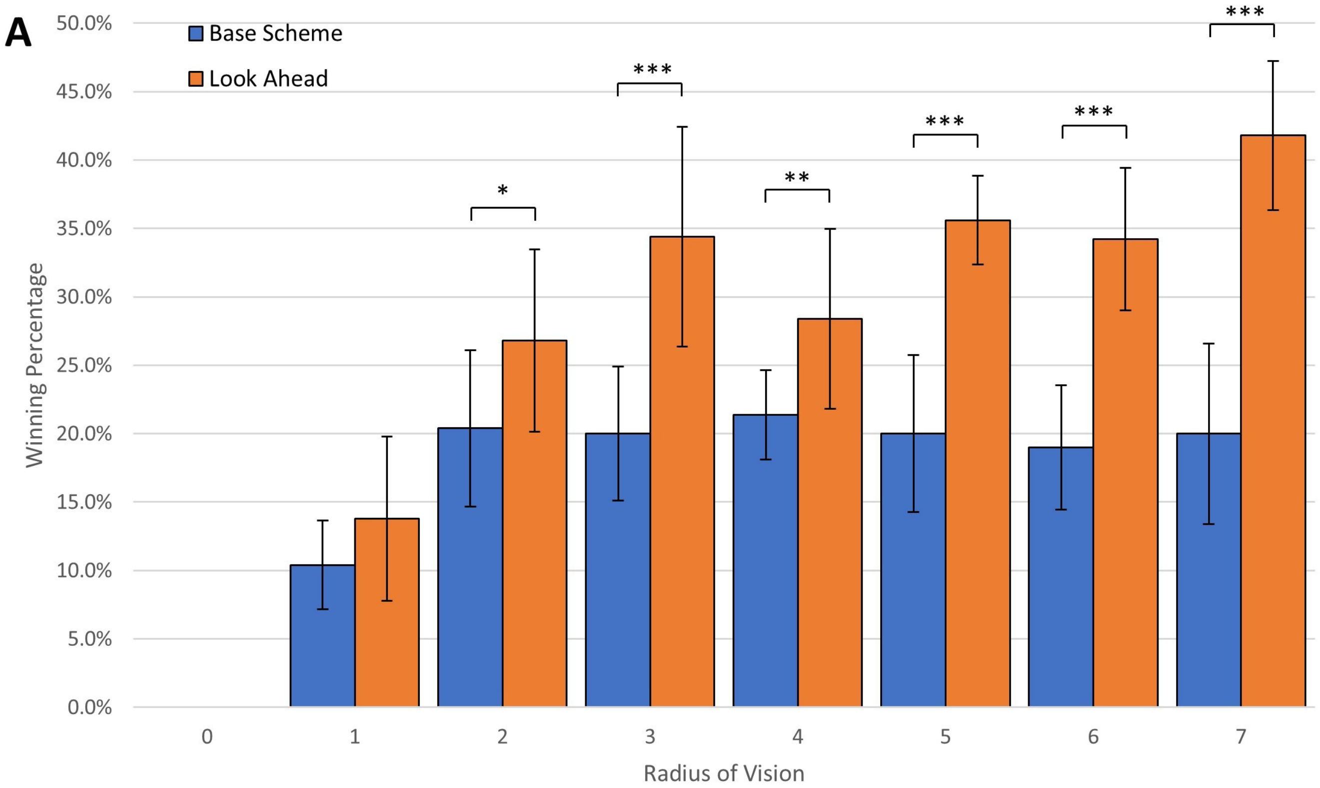


Figure 8: Performance is significantly better when pieces look to attack one move ahead, compared to the base scheme. A radius-by-radius performance comparison between the base scheme, and the base scheme with pieces looking one move ahead (N=50).

References

Adamatzky, A, Chiolerio, A, & Szacilowski, K 2020 “Liquid metal droplet solves maze”, *Soft Matter*, vol. 16, no. 6, pp. 1455–1462.

Adami, C 2002 “What is complexity?”, *Bioessays*, vol. 24, no. 12, pp. 1085–1094.

Albantakis, L, *et al.* 2014 “Evolution of integrated causal structures in animats exposed to environments of increasing complexity”, *PLoS Computational Biology*, vol. 10, no. 12 article e1003966.

Albantakis, L, *et al.* 2017 “What caused what? An irreducible account of actual causation”, *arXiv*, article arXiv:1708.06716.

Amidzic, O, *et al.* 2001 “Pattern of focal γ-bursts in chess players”, *Nature*, vol. 412, no. 6847, pp. 603–603.

Badcock, PB, *et al.* 2019 “The hierarchically mechanistic mind: An evolutionary systems theory of the human brain, cognition, and behavior”, *Cognitive, Affective, & Behavioral Neuroscience*, vol. 19, no. 6, pp. 1319–1351.

Baluška, F, & Levin, M 2016 “On having no head: Cognition throughout biological systems”, *Frontiers in Psychology*, vol. 7 https://doi.org/10.3389/fpsyg.2016.00902

Barca, JC, & Sekercioglu, YA 2013 “Swarm robotics reviewed”, *Robotica*, vol. 31: pp. 345–359.

Bazazi, S, *et al.* 2008 “Collective motion and cannibalism in locust migratory bands”, *Current Biology*, vol. 18, no. 10, pp. 735–739.

Beer, RD 2004 “Autopoiesis and cognition in the game of life”, *Artificial Life*, vol. 10, no. 3, pp. 309–326.

Beer, RD 2014 “The cognitive domain of a glider in the game of life”, *Artificial Life*, vol. 20, no. 2, pp. 183–206.

Berdahl, AM, *et al.* 2018 “Collective animal navigation and migratory culture: From theoretical models to empirical evidence”, *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 373, no. 1746.

Bilalić, M, McLeod, B, & Gobet F 2007 "Does chess need intelligence? A study with young chess players", *Intelligence*, vol. 35, no. 4, pp. 457–470.

Blackiston, D, *et al.* 2023 "Biological robots: Perspectives on an emerging interdisciplinary field", *Soft Robot*, vol. 10, no. 4, pp. 674–686. DOI: 10.1089/soro.2022.0142

Boddy, AM, *et al.* 2015 "Fetal microchimerism and maternal health: A review and evolutionary analysis of cooperation and conflict beyond the womb", *Bioessays*, vol. 37, no. 10,pp. 1106–1118.

Bongard, J, & Levin, M 2023 "There's plenty of room right here: Biological systems as evolved, overloaded, multi-scale machines", *Biomimetics* (Basel), vol. 8, no. 110. https://doi.org/10.3390/biomimetics8010110

Brambilla, M, *et al.* 2013 "Swarm robotics: A review from the swarm engineering perspective", *Swarm Intelligence*, vol. 7, no. 1, pp. 1–41.

Braude, SE 1995 *First Person Plural: Multiple Personality and the Philosophy of Mind*. Lanham, MD: Rowman & Littlefield Publishers.

Bullmore, E, & Sporns, O 2009 "Complex brain networks: Graph theoretical analysis of structural and functional systems", *Nature Reviews. Neuroscience*, vol. 10, no. 3, pp. 186–198.

Campitelli, G, & Gobet, F 2008 "The role of practice in chess: A longitudinal study", *Learning and Individual Differences*, vol. 18, no. 4, pp. 446–458.

Campitelli, G, *et al.* 2007 "Brain localization of memory chunks in chessplayers", *International Journal of Neuroscience*, vol. 17, no. 12, pp. 1641–1659.

Campitelli, G, Gobet, F, & Bilalić, M 2014 "Cognitive processes and development of chess genius: An integrative approach". In: Simonton, DK (ed.) *The Wiley Handbook of Genius*, pp. 350–374. New York: John Wiley & Sons, Ltd.

Capablanca, JRI 1921 *Chess Fundamentals*. New York: Harcourt, Brace and company.

Capablanca, JR 2006 *Chess Fundamentals*. New York, NY: Random House.

Cejkova, J, *et al.* 2017 "Droplets as liquid robots", *Artificial Life*, vol. 23, no. 4, pp. 528–549.

Chater, N 2018 *The Mind is Flat: The Illusion of Mental Depth and the Improvised Mind*. London: Allen Lane.

Chess.com, 2024a *Analysis Board* [Online tool]. Retrieved from https://www.chess.com/analysis, 2024, March 26.

Chess.com, 2024b *Chess Terms* [Online tool]. Retrieved from https://www.chess.com/terms, 2024, March 26.

Couzin, I 2007 "Collective minds", *Nature*, vol. 445, no. 715 https://doi.org/10.1038/445715a.

Couzin, ID 2009 "Collective cognition in animal groups", *Trends in Cognitive Sciences*, vol. 13, no. 1, pp. 36–43.

Couzin, ID 2018 "Synchronization: The key to effective communication in animal collectives", *Trends in Cognitive Sciences*, vol. 22, no. 10, pp. 844–846.

Couzin, ID, *et al.* 2002 "Collective memory and spatial sorting in animal groups", *Journal of Theoretical Biology*, vol. 218, no. 1, pp. 1–11.

Dangauthier, P, *et al.* 2007 "TrueSkill through time: Revisiting the history of chess", *Advances in Neural Information Processing Systems 20, Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 3–6.*

Davidson, HA 2012 *A Short History of Chess*. New York: Crown.

Deisboeck, TS, & Couzin, ID 2009 "Collective behavior in cancer cell populations", *Bioessays*, vol. 31, no. 2, pp. 190–197.

Dewan, EM, 1976 "Consciousness as an emergent causal agent in the context of control system theory". In: Globus, GMG, & Savodnik, I (eds.) *Consciousness and the Brain: A Scientific and Philosophical Inquiry*, pp. 181–198. New York: Plenum Press.

Duca Iliescu, DM 2020 "The impact of artificial intelligence on the chess world", *JMIR Serious Games*, vol. 8, no. 4, article e24049.

Edlund, JA, *et al.*, 2011 "Integrated information increases with fitness in the evolution of animats", *PLoS Computational Biology*, vol. 7, no. 10, article e1002236.

Elo, AE 2008 *The Rating of Chessplayers: Past and Present*. Bronx, NY: Ishi Press International.

FIDE (International Chess Federation), *FIDE Laws of Chess*. Retrieved from https://handbook.fide.com/chapter/E012023, 2023, January 1.

Fields, C, & Levin, M 2022 "Competency in navigating arbitrary spaces as an invariant for analyzing cognition in diverse embodiments", *Entropy* (Basel), vol. 24, no. 6.

Friston, K 2013 "Life as we know it", *Journal of the Royal Society Interface*, vol. 10, no. 86, article 20130475.

Friston, K, & Buzsaki, G 2016 "The functional anatomy of time: What and when in the brain", *Trends in Cognitive Sciences*, vol. 20, no. 7, pp. 500–511.

Friston, K, & Frith, C 2015 "A Duet for one", *Consciousness and Cognition*, vol. 36, pp. 390–405. doi: 10.1016/j.concog.2014.12.003.

Fujita, K 2022 "AlphaDDA: Strategies for adjusting the playing strength of a fully trained AlphaZero system to a suitable human training partner", *PeerJ Computer Science*, vol. 8, article e1123.

Gad, AF 2023 "PyGAD: An intuitive genetic algorithm Python Library", *Multimedia Tools and Applications*, vol. 83, pp 58029–58042. https://doi.org/10.1007/s11042-023-17167-y.

Gawne, R, McKenna, KZ, & Levin, M 2020 "Competitive and coordinative interactions between body parts produce adaptive developmental outcomes", *Bioessays*, vol. 42, no. 8, article e1900245.

Gazzaniga, MS 2005 “Forty-five years of split-brain research and still going strong”, *Nature Reviews Neuroscience*, vol. 6, no. 8, pp. 653–659.

Gomes, J, Urbano, P, & Christensen, AL 2013 “Evolution of swarm robotics systems with novelty search”, *Swarm Intelligence*, vol. 7, no. 2-3, pp. 115–144.

Gordon, DM 2016a “Collective wisdom of ants”, *Scientific American*, vol. 314, no. 2, pp. 44–47.

Gordon, DM 2016b “From division of labor to the collective behavior of social insects”, *Behavioral Ecology and Sociobiology*, vol. 70, pp. 1101–1108.

Gordon, DM 2016c “The evolution of the algorithms for collective behavior”, *Cell Systems*, vol. 3, no. 6, pp. 514–520.

Griffin, WA 2006 “Agent-based modeling for the theoretical biologist”, *Biological Theory*, vol. 1, no. 4, pp. 404–409.

Hanczyc, MM 2014 “Droplets: Unconventional protocell model with life-like dynamics and room to grow”, *Life* (Basel), vol. 4, no. 4, pp. 1038–1049.

Hearst, E, & Knott, J 2009 *Blindfold Chess: History, Psychology, Techniques, Champions, World Records, and Important Games*. Jefferson, NC: McFarland.

Heider, F, & Simmel, M 1944 “An experimental study of apparent behavior”, *The American Journal of Psychology*, vol. 57, pp. 243–259. https://doi.org/10.2307/1416950

Heylighen, F 2013 “Self-organization in communicating groups: The emergence of coordination, shared references and collective intelligence”, pp. 117–149. In: Massip-Bonet, À, & Bastardas-Boada, A (eds.) *Complexity Perspectives on Language, Communication and Society. Understanding Complex Systems*. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-32817-6_10.

Hoel, EP, & *et al.* 2016 “Can the macro beat the micro? Integrated information across spatiotemporal scales”, *Neuroscience of Consciousness*, vol. 2016, article niw012.

Hoel, EP, Albantakis, L, & Tononi, G 2013 “Quantifying causal emergence shows that macro can beat micro”, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 49, pp. 19790–19795.

James, W 1890 *The Principles of Psychology. American Science—Advanced Course*. New York,: H. Holt and company.

K, B., 2021 “AI in chess: The evolution of artificial intelligence in chess engines”, *Medium*, 2 April.

Kolchinsky, A, *et al.* 2014 “Multi-scale integration and predictability in resting state brain activity”, *Frontiers in Neuroinformatics*, vol. 8, p. 66.

Lagasse, E, & Levin, M 2023 “Future medicine: From molecular pathways to the collective intelligence of the body”, *Trends in Molecular Medicine*, vol. 29, no. 9, pp. 687–710.

Letendre, K, & Moses, ME 2013 “Synergy in ant foraging strategies: Memory and communication alone and in combination”. In: *Gecco’13: Proceedings of the 2013 Genetic and Evolutionary Computation Conference*, pp. 41–48.

Levin, M 2019 “The computational boundary of a ‘Self’: Developmental bioelectricity drives multicellularity and scale-free cognition”, *Frontiers in Psychology*, vol. 10, no. 2688, article 2688.

Levin, M 2021 “Bioelectrical approaches to cancer as a problem of the scaling of the cellular self”, *Progress in Biophysics and Molecular Biology*, vol. 165, pp. 102–113.

Levin, M 2022 “Technological approach to mind everywhere: An experimentally-grounded framework for understanding diverse bodies and minds”, *Frontiers in Systems Neuroscience*, vol. 16, article 768201.

Levin, M 2023a “Collective intelligence of morphogenesis as a teleonomic process”, pp. 175–198. In: Corning, PA, Kauffman, SA, Noble, D, Shapiro, JA, Vane-Wright, RI, & Pross, A (eds.) *Evolution “on Purpose”: Teleonomy in Living Systems*. Cambridge: MIT Press.

Levin, M 2023b “Bioelectric networks: The cognitive glue enabling evolutionary scaling from physiology to mind”, *Animal Cognition*, vol. 26, pp. 1865–1891.

Levin, M 2023c “Darwin’s agential materials: Evolutionary implications of multiscale competency in developmental biology”, *Cellular and Molecular Life Sciences*, vol. 80, no. 6, article 142.

Levin, M, *et al.* 2021 “Uncovering cognitive similarities and differences, conservation and innovation”, *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 376, no. 1821, article 20200458.

Levin, M 2024 “Self-improving memories: A perspective on memories as agential, dynamically-reinterpreting cognitive glue”, *Entropy*, vol. 26, no. 481.

Liang, D, *et al.* 2022 “Training-specific changes in regional spontaneous neural activity among professional Chinese chess players”, *Frontiers in Neuroscience*, vol. 16, article 877103.

Lyon, P 2015 “The cognitive cell: Bacterial behavior reconsidered”, *Frontiers in Microbiology*, vol. 6, no. 264.

Lyon, P 2006 “The biogenic approach to cognition”, *Cognitive Processing*, vol. 7, no. 1, pp. 11–29.

Lyon, P 2020 “Of what is ‘minimal cognition’ the half-baked version?”, *Adaptive Behavior*, vol. 28, no. 6, pp. 407–424.

Lyon, P, *et al.* 2021 “Reframing cognition: Getting down to biological basics”, *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 376, no. 1820, article 20190750.

Maharaj, S, Polson, N, & Turk, A 2022 “Chess AI: Competing paradigms for machine intelligence”, *Entropy* (Basel), vol. 24, no. 4.

Manicka, S, & Levin, M 2019 “The cognitive lens: A primer on conceptual tools for analysing information processing in developmental and regenerative morphogenesis”, *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 374, no. 1774, article 20180369.

Martinez-Corral, R, *et al.* 2019 “Metabolic basis of brain-like electrical signalling in bacterial communities”, *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 374, no. 1774, article 20180382.

Mathews, J, *et al.* 2023 "Cellular signaling pathways as plastic, proto-cognitive systems: Implications for biomedicine", *Patterns* (N Y), vol. 4, no. 5, article 100737.

McMillen, P, & Levin, M 2024 "Collective intelligence: A unifying concept for integrating biology across scales and substrates", *Communications Biology*, vol. 7, no. 1, article 378.

Miller, SD, & Triggiano, PJ 1992 "The psychophysiological investigation of multiple personality disorder: Review and update", *The American Journal of Clinical Hypnosis*, vol. 35, no. 1, pp. 47–61.

Montgomery, BA 2003 *Consciousness and Personhood in Split-Brain Patients*. Dissertation, University of Oklahoma. https://philpapers.org/rec/MONCAP-3.

Paulson, S, *et al.* 2017 "The deeper self: An expanded view of consciousness", *Annals of the New York Academy of Sciences*, vol. 1406, no. 1, pp. 46–63. DOI: 10.1111/nyas.13403.

Pedraja, F, & Sawtell, NB 2024 "Collective sensing in electric fish", *Nature*, vol. 628, no. 8006, pp. 139–144.

Pezzulo, G, Rigoli, F, & Friston, K 2015 "Active inference, homeostatic regulation and adaptive behavioural control", *Progress in Neurobiology*, vol. 134, pp. 17–35.

Pinero, J, & Sole, R 2019 "Statistical physics of liquid brains", *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 374, no. 1774, article 20180376.

Pio-Lopez, L, *et al.* 2023 "The scaling of goals from cellular to anatomical homeostasis: An evolutionary simulation, experiment and analysis", *Interface Focus*, vol. 13, no. 3, article 20220072.

Prokopenko, M, Boschietti, F, & Ryan, AJ 2009 "An information-theoretic primer on complexity, self-organization, and emergence", *Complexity*, vol. 15, no. 1, pp. 11–28.

Putnam, FW 1992 "Multiple personality disorder", *British Journal of Psychiatry*, vol. 161, pp. 415–416 (Author reply: pp. 418–420).

Rahwan, I, *et al.* 2019. "Machine behaviour", *Nature*, vol. 568, no. 7753, pp. 477–486.

Ramstead, MJD, *et al.* 2019 "Variational ecology and the physics of sentient systems", *Physics of Life Reviews*, vol. 31, pp. 188–205.

Ramstead, MJ, *et al.* 2022 "On Bayesian mechanics: A physics of and by beliefs", *ArXiv preprint* arXiv:2205.11543.

Reid, C.R., *et al.* 2015a "Army ants dynamically adjust living bridges in response to a cost-benefit trade-off", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 49, pp. 15113–15118.

Reid, CR, *et al.* 2015b. "Information integration and multiattribute decision making in non-neuronal organisms", *Animal Behaviour*, vol. 100, pp. 44–50.

Reid, CR, *et al.* 2016 "Decision-making without a brain: How an amoeboid organism solves the two-armed bandit", *Journal of the Royal Society Interface*, vol. 13, no. 119.

Rozman, L 2023 *How to Win at Chess: The Ultimate Guide for Beginners and Beyond*. New York: Ten Speed Press.

Schmid, M, *et al.* 2023 "Student of games: A unified learning algorithm for both perfect and imperfect information games", *Science Advances*, vol. 9, no. 46, article eadg3256.

Scholl, BJ, & Tremoulet, PD 2000 "Perceptual causality and animacy", *Trends in Cognitive Sciences*, vol. 4, no. 8, pp. 299–309.

Seoane, LF 2019 "Evolutionary aspects of reservoir computing", *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 374, no. 1774, article 20180377.

Seth, AK 2013 "Interoceptive inference, emotion, and the embodied self", *Trends in Cognitive Sciences*, vol. 17, no. 11, pp. 565–573.

Sharples, J 2017 *A Cultural History of Chess-players: Minds, Machines, and Monsters*. Manchester: Manchester University Press.

Shreesha, L, & Levin, M 2024 "Stress sharing as cognitive glue for collective intelligences: A computational model of stress as a coordinator for morphogenesis", *Biochemical and Biophysical Research Communications*, vol. 731, article 150396.

Sole, R, *et al.* 2016, "Synthetic collective intelligence", *Biosystems*, vol. 148, pp. 47–61.

Sole, R, Moses, M, & Forrest, S 2019 "Liquid brains, solid brains", *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 374, no. 1774, article 20190040.

Solms, M 2018 "The hard problem of consciousness and the free energy principle", *Frontiers in Psychology*, vol. 9, article 2714.

Song, L, *et al.* 2022 "Professional chess expertise modulates whole brain functional connectivity pattern homogeneity and couplings", *Brain Imaging and Behavior*, vol. 16, no. 2, pp. 587–595.

Sporns, O 2011 *Networks of the Brain*. Cambridge, Mass.: MIT Press.

Steinbacher, M, *et al.* 2021 "Advances in the agent-based modeling of economic and social behavior", *SN Business & Economics*, vol. 1, no. 7, article 99. doi: 10.1007/s43546-021-00103-3

Strassmann, JE, & Queller, DC 2010 "The social organism: Congresses, parties, and committees", *Evolution*, vol. 64, no. 3, pp. 605–616.

Strong, V, Holderbaum, W, & Hayashi, Y 2024 "Electro-active polymer hydrogels exhibit emergent memory when embodied in a simulated game environment", *Cell Reports Physical Science*, vol. 5, no. 9.

The Stockfish developers, n.d. Stockfish [Computer software]. https://github.com/official-stockfish/Stockfish.

Tononi, G, & Sporns, O 2003. "Measuring information integration", *BMC Neuroscience*, vol. 4, article 31.

Tononi, G, Edelman, GM, & Sporns, O 1998 "Complexity and coherency: Integrating information in the brain", *Trends in Cognitive Sciences*, vol. 2, no. 12, pp. 474–484.

Tononi, G, Sporns, O, & Edelman, GM 1994 "A measure for brain complexity: Relating functional segregation and integration in the nervous system", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, no. 11, pp. 5033–5037.

Trianni, V, & Campo, A 2015 "Fundamental collective behaviors in swarm robotics, pp. 1377–1394. In: Kacprzyk, J, & Pedrycz, W (eds.) *Springer Handbook of Computational Intelligence*. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-43505-2_71

Vallverdu, J, *et al.* 2018 "Slime mould: The fundamental mechanisms of biological cognition", *Biosystems*, vol. 165, pp. 57–70.

Ward, AJ, *et al.* 2008 "Quorum decision-making facilitates information transfer in fish shoals", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 19, pp. 6948–6953.

Weng, J, *et al.* 2001 "Artificial intelligence. Autonomous mental development by robots and animals", *Science*, vol. 291, no. 5504, pp. 599–600.

Wheeler, WM 1911 "The ant-colony as an organism", *Journal of Morphology*, vol. 22, no. 2, pp. 307–325.

Witkowski, O, & Ikegami, T 2019 "How to make swarms open-ended? Evolving collective intelligence through a constricted exploration of adjacent possibles", *Artificial Life*, vol. 25, no. 2, pp. 178–197.

Zhang, T, Goldstein, A, & Levin, M 2024 "Classical sorting algorithms as a model of morphogenesis: Self-sorting arrays reveal unexpected competencies in a minimal model of basal intelligence", *Adaptive Behavior*, vol. 0, no. 0, article 10597123241269740.

Zhelyabuzhsky, I 2022 Stockfish 3.28.0 [Python Library], PyPI. https://pypi.org/project/stockfish/.

Organisms

Special Issue, “What AI Can Learn from Biology”

Vol. 8, No. 1–2 (2025)
ISSN: 2532-5876
Open access journal licensed under CC-BY
DOI: 10.13133/2532-5876/19211

What Drives the Brain?
Organizational Changes, FEP and Anti-entropy

Marie Chollat-Namy[a] *and Maël Montévil*[a*]

[a]*Centre Cavaillès République des Savoirs, École Normale Supérieure, Paris, France*

***Corresponding author:** Maël Montévil, Email: mael.montevil@ens.psl.eu

Abstract

The free-energy principle (FEP) provides a computational, physical and teleological theory for understanding biological organization as cognitive agent minimizing their entropy in relation to their environment. Is minimizing entropy the first principle driving all dynamics of cognition? Is it enough to account for organizational changes in an open-ended way? After a general presentation of the literature on the FEP, we turn to the paradoxical case of the brain under the influence of psychedelics, where the FEP is challenged by an increased cerebral entropy, which induces organizational changes of the cognition. Building on this paradox, we identify some limits of the FEP, notably applying concepts of information, optimization and predefined phase space to biology that do not fit our criteria for a theory of biological organization. We also identify two aspects of entropy in physics and in the FEP: the local entropic trend that implies variations and the global entropic trend that leads to homogeneization and stability. Extending these concepts outside of their physicomathematical context, we contribute to an organicist theoretical alternative where living systems find a balance between these two trends, and, conceptually, a biological system’s disorganization enable its “unprestatable” reorganization and so its open-ended evolution.

Keywords: free energy principle, anti-entropy, entropy, novelty, biological organization, organicism

Citation: Chollat-Namy, M, & Montévil, M 2025, “What Drives the Brain? Organizational Changes, FEP and Anti-entropy”, *Organisms: Journal of Biological Sciences*, vol. 8, no. 1–2, pp. 63–84. DOI: 10.13133/2532-5876/19211

Introduction

Scientists often mobilize approaches from fields other than biology to understand living beings. However, it is not easy in biology to apply, for example, the principles of physics, which are based on conservation, optimization, and the pre-definition of the phase space. Indeed, as René Thom put it, "it is the lack of the definition [of the virtual possible] that affects – very seriously – the scientific nature of Darwin's Theory of Evolution" (Amsterdamski 1990). Similarly, since the discovery of the physicochemical structure of DNA as the material support of genes, principles derived from computer science have been widely applied to understand living organisms with a reductionist, genocentric viewpoint. They remain so today, even if their validity has been belied by numerous discoveries and analyses, such as the diversity of gene reading modes, alternative splicing, epigenetics, and developmental plasticity.

In this context, an interdisciplinary effort aims to rework the conceptual framework for understanding biological organisms by following an organicist approach that is neither physicalist nor "informationalist". This work has led to proposing three principles for a theory of organisms. The principle of variation posits the historicity of biological objects: the regularities of living beings playing a causal role, which we call constraints, are part of a history and can change over time. Biological objects cannot be defined based on invariants and symmetry as in physics; we say they are specific (Montévil *et al.* 2016a). The practical way of defining them is phylogenetics, sometimes also genealogy for laboratory strains, but in all cases, it is historical (Montévil 2019). If these objects are initially variable, the relative stability of their constraints needs to be explained. The principle of organization has this function: in an organism, a constraint canalizes a process that maintains another constraint, which canalizes a process, and so on, leading to circularity called the closure of constraints (Mossio, Montévil & Longo 2016; Montévil & Mossio 2015). This principle also allows us to speak of function in the sense of the relationship between a part and a whole, defined by the circularity of the closure. Finally, we posit that the default state of cells, i.e., their behavior when no particular cause acts upon them, is proliferation and motility, not quiescence (Soto, Longo, Montévil *et al.* 2016). In other words, living beings do not need stimulation to be active.

In line with this framework, anti-entropy has been introduced as an addition to entropy. The term anti-entropy stems from an analogy with anti-matter: anti-matter is symmetrical to matter, but has opposite properties in some respects. Anti-entropy was first introduced as a measure of phenotypic complexity and addressed through its metabolic consequences (Bailly & Longo 2009). The idea has since been refined based on biological variations interpreted as changes in symmetries, i.e., what was to become the principle of variation (Longo & Montévil 2012). Finally, the most recent concept defines the production of anti-entropy, by analogy with the production of entropy, as the production of a functional novelty, i.e., the production of a singular situation that contributes to the closure of an organization by this singularity (Montévil 2021). Indeed, entropy production provides the arrow of time of physics by the second principle of thermodynamics and the idea that a system spontaneously moves from being somewhat specific to the most generic configuration, given the constraints of that system. Biological historicity manifests a second time arrow, with objects that can produce increasingly functional specificity.

In this context, many questions remain. For example, closure of constraints does not imply that an organization remains unchanged. On the contrary, the principle of variation means that biological organizations can always change, but how do these changes take place? In particular, what is the relationship between organizational change and entropy? In the particular case where these changes correspond to functional novelties, i.e., correspond to a production of anti-entropy, what is their relationship to entropy?

These questions are met by an informational approach to biology defining the "free energy principle" (FEP). Schrödinger, in his book What is Life? (Schrödinger 1944), proposed that the primary characteristic of living systems is repelling entropy while maintaining their internal order. Similarly, the info-computational vision of the FEP understands the organization of living systems as the result of a computational process based on the minimization of entropy. The FEP aims to provide a mathematical framework for the temporal evolution of a living system and that of its model of "beliefs" in terms of Bayesian updates optimized to fit the statistics of the things to which the system is coupled (Ramstead *et al.* 2023). This theory states that living systems seek to minimize the variational free energy corresponding to the relative entropy of the system's generative model. This info-computational approach provides a self-organizing model of the living world, where organisms are made of layers of nested abstract representations generating probabilistic decisions (Kirchhoff *et al.* 2018).

The FEP is mainly used in neuroscience to formalize the leading theory in this field, namely the Bayesian brain theory. According to this theory, the brain actively infers the causes of its sensations and selects actions to minimize entropy relative to its subject. Thus, the Bayesian generative model of the brain updates and evolves by maximizing the evidence for its beliefs (Friston, Kilner & Harrison 2006). The FEP is a variational principle, posited as equivalent to the principle of least action, fundamental in physics, and the principle of maximum entropy, but applying to Bayesian mechanics as a "physics of and by beliefs" (Ramstead *et al.* 2023). This informational principle states that living systems tend to optimal maintenance and adaptation to their environment by organizing themselves against entropy. According to proponents of this theory, it applies to all living beings, even those without nervous systems, and even to all evolutionary phenomena, biological (Kuchling *et al.* 2020; Kirchhoff *et al.* 2018; Campbell 2016) and societal (Slijepcevic 2024). Thus, according to some of its advocates, it is a candidate for "a great unifying theory" (Sánchez-Cañizares 2021).

As argued in the first section, the FEP is incompatible *ex hypothesi* with the organicist framework we are helping to develop, as it develops an informational approach to living things and lean on optimization principles (Bailly & Longo 2011; Longo *et al.* 2012). Nevertheless, the FEP and its critique allow us to work on the relationship between entropy, organization and changes in organization (Chollat-Namy & Longo 2022). We begin with a general presentation of the literature on FEP as a principle of cognition and organization at all levels of living organisms. We then turn to the paradoxical and much-discussed case of the brain under the influence of psychedelics. This case is interesting since the FEP organizing principle is challenged by an increased cerebral entropy, which nonetheless seems to induce beneficial changes at both neurological and psychological levels.

Building on this paradox, we will criticize the FEP, first pointing out some general difficulties in applying information concepts in biology, then more specifically on the physicalist assumptions of the FEP, notably the existence of a predefined phase space. The aim will not be to reject all the ideas put forward by FEP theorists but to demonstrate some of their limits and contribute to overcoming them by proposing an organicist theoretical alternative based on current work in this field.

By analyzing entropy within living systems, we will add to the concept of anti-entropy, explaining how a biological system's disorganization can enable its reorganization and evolution towards new, viable, and not only unpredictable but also "unprestatable" configurations; that is, the changes are not just about a state among predefined possible states, but the possibilities themselves are unpredictable. This approach will lead us to rediscuss the default state of life and the notions of causality and finality in biology outside a physicalist paradigm.

1. Informational Theory of Cognition and Entropy Minimization as a Theoretical Principle

1.1. The Principle of Free Energy and its Application to Biological Organization

Many researchers have argued that algorithmic information processing by living systems is essential to their stability and survival (Walker & Davies 2013), and involves capturing information about their environment, then translating this information into exploitable and adaptive actions. It has been suggested that this process is the defining characteristic of living organisms and would be uniquely oriented towards maintaining organisms in their expected phenotypic and ontogenetic state (Kirchhoff *et al.* 2018).

This maintenance objective is achieved by the free energy principle, according to which living systems seek to minimize a theoretical quantity of information called "free energy," corresponding to the entropy relative to the system and its coupling to its environment. According to this theory, any biological organization, in particular the nervous system, creates statistical approximations, Bayesian generative models, corresponding to a hierarchical system of "beliefs" about the causes of its sensory data (Knill & Pouget 2004; Friston, Kilner & Harrison 2006; Friston & Kiebel 2009). A system minimizes its free energy when it implicitly optimizes its "belief" about what provokes sensory input. In other words, a living system actively infers and projects hypotheses about the causes of its sensations and selects actions to minimize the relative entropy about them. This free energy is also called "uncertainty," "surprise," or "prediction error," and minimizing it is equivalent to maximizing the evidence for the belief model (Kiverstein, Kirchhoff & Froese 2022). This localized control of entropy would act as a "driving force" for the adaptive reconciliation of living systems with their environment and thus towards their stability.

In this sense, the FEP implies that all living systems, considered to be endowed with cognition, can be modeled as visiting a limited set of states in order to continue to exist (Parr & Friston 2019). This modeling uses information geometry techniques that formally specify the boundary between a living system and its external environment, notably as a Markov blanket (Palacios *et al.* 2020). A Markov blanket is based on a statistical partitioning between internal states (systemic) and external (environmental) states. The Markov blanket includes a second partitioning between active and passive sensory states, mediating exchanges between internal and external states (Ramstead *et al.* 2021).

This info-computational and cognitivist vision of the living world, based on Bayesian model generation through FEP action, is applied beyond the brain (Kirchhoff *et al.* 2018; Slijepcevic 2024) and could be used to describe any type of biological evolutionary phenomenon, including morphological development (Kuchling *et al.* 2020), phylogenetic evolution, psychology and even the evolution of societies and scientific knowledge (Campbell 2016). In the case of phylogenetic evolution, for example, the set of "instructions" for growth and development that an organism inherits constitutes a kind of prediction about the organism's suitability for its environment. It is as if a phenotype were actively inferring the state of its ecosystemic niche under a generative model, whose parameters are learned through natural selection, seen as the optimization process of the Bayesian model (Friston *et al.* 2023; Czégel *et al.* 2022).

This theory considers that living organisms and their various forms are organized according to a generative computational model oriented towards their maintenance and adaptation to the environment by the FEP. It is mainly used in neuroscience to understand cognition's adaptive and learning capacities (Friston, Kilner & Harrison 2006).

1.2. The Principle of Free Energy Challenged by the Brain under Psychedelics

The FEP has been heavily discussed in the particular case of the brain under the influence of psychedelics. This case is interesting because it is challenging the FEP. The brain exhibits an increased entropy, which seems beneficial for inducing biologically novel and psychologically therapeutic changes. However, the FEP considers that cognitive systems must constantly minimize their entropy relative to their coupling to their environment, which correspond to "the long-term average of surprise", defined as "the difference between an organism's predictions about its sensory inputs (embodied in its models of the world) and the sensations it actually encounters." (Friston *et al.* 2012). This principle apparently contradicts the phenomena observed during the psychedelic experience. The theory of the entropic brain and its new version, REBUS (RElaxed Beliefs Under pSychedelics) (Carhart-Harris & Friston 2019), aims to overcome this paradox.

Psychedelics, including LSD, psilocybin, DMT, mescaline and many others, are natural or synthetic substances that act on the brain's serotonin network, producing intense psychological and physiological effects. Legal restrictions have limited their use in clinical research for several decades. However, in recent years, these substances have become the subject of active research, and numerous studies have revealed the therapeutic potential of these molecules to treat a variety of psychological problems, such as addiction (Zafar *et al.* 2023), end-of-life anxiety (Whinkin *et al.* 2023), post-traumatic syndromes (Fonseka & Woo 2023) and depression (Hristova & Pérez-Jover 2023; Rivera-García & Cruz 2023). Although their molecular mechanisms of action, through interaction with serotonin receptors, notably 5HT1A and 2A, are well known (Cameron *et al.* 2023), they are not sufficient to explain their effect on the dynamic organization of the brain and psyche, which requires a specific theoretical approach. The leading theory today is that of J. Carhart Harris, known as REBUS (Carhart-Harris & Friston 2019). It is based on the principle that thanks to their entropic effect on spontaneous cortical activity, psychedelics act to relax the precision of high-level hierarchical beliefs, freeing up activity at lower levels (note that this effect is especially visible at high doses of psychedelics, and that lower doses may induce the opposite effect, a strengthening of beliefs (Safron *et* *al*. 2020)). This theory mobilizes the principle of free energy. As mentioned above, the FEP describes brain behavior based on its inherent tendency to resist disorder and minimize uncertainty by optimizing, through Bayesian updating, its probabilistic representations and sampling of its environment. These representations, or *a priori* beliefs, constitute predictive processing organized in hierarchical levels (Friston 2010).

In the Bayesian vision of the brain, bottom-up sensory input is compared with inferred top-down predictions. The resulting prediction error is then passed on to higher hierarchies to update the representations, generating top-down predictions on lower levels (Badcock *et al*. 2019). Following the FEP, Neural dynamics attempts to minimize the amplitude of prediction errors at each hierarchy level. This process provides an optimized causal explanation of sensory input at several levels of hierarchical abstraction. The highest levels form compressive synthesis from the content of the lower levels they envelop, thus reducing their potential information content (Ruffini 2017).

The apex of this hierarchy of prior belief levels is instantiated by the DMN, the "default mode network" (Margulies *et al*. 2016; Carhart-Harris and Friston 2019) considered to be the seat of the sense of self, of identity as "internal narrative" (Menon 2023). The DMN, functionally positioned as far as possible from sensorimotor input (Smallwood *et al*. 2021), is associated with subjective states such as reflection, remembering, introspection, planning, social interaction, abstract thought... (Buckner, Andrews-Hanna & Schacter 2008; Menon 2023).

The theory's central idea is that psychedelics increase the entropy, the variational free energy, of brain activity and reduce the precision (inverse variance, felt confidence, rigidity) of higher-level prior beliefs, making them more sensitive to bottom-up prediction errors. This process would disrupt the DMN's directional function and relax prior beliefs, making them more plastic and susceptible to change into new configurations of meaning. Thus, at the psychological level, psychedelic intake can temporarily induce a feeling of uncertainty but also intuitive understandings and changes in perspective about oneself and the world (Timmermann *et al*. 2021). At the biological level, we observe that brain dynamics display increased complexity, the construction of new and diverse connectivities is promoted (Carhart-Harris 2018), and neurogenesis is stimulated (Calder & Hasler 2023).

In short, taking psychedelics in significant quantities "disorganizes" the cerebral hierarchy temporarily and seems to induce the creation of new configurations, sometimes functional at the physiological level and meaningful at the psychological level. Why?

Although psychedelics appear to "temporarily breach the free energy principle" (Carhart-Harris & Friston 2019), the authors point to a higher level at which free energy would be minimized, inducing a revision of beliefs about generative models themselves. This process would be achieved by selecting the best-performing model from a set of models (Bayesian Model Selection) or reducing complexity (Bayesian Model Reduction) by removing redundant model parameters. These mechanisms for simplifying and generalizing the model would produce "inferences used to fill an explanatory gap." This type of inference would underlie the experience of insight (also called "eureka" moment or intuitive understanding) (Friston *et al*. 2017) and explain the changes in point of view generated by the psychedelic experience (Carhart-Harris & Friston 2019).

These mechanisms would also be responsible for recalibrating the relevant beliefs to be better functionally harmonized with the other levels. As the cause of many psychological illnesses is the pathological weighting of certain prior beliefs, this process of recalibrating beliefs could explain the therapeutic effect of psychedelics on mental health over the long term (Carhart-Harris & Friston 2019).

On the informational level, the effect of psychedelics can be modeled as a reduction of the curvature of the energetic landscape that contains neuronal dynamics and a flattening of the local minima. This phenomenon allows neuronal dynamics to escape its attractor's basins and prior beliefs and explore the space of state with fewer constraints. The authors consider this flattening of the energy landscape of the brain by psychedelics as analogous to the method of annealing in computer science, a method of optimization to find new local minima. Inspired by metallurgy, this approach is performed in two steps. First, the system is "heated". It reaches a state of increased plasticity to discover "new" relatively stable low-energy states where the system can reside at lower temperatures (Wang & Smith 1998). This method is also similar to new approaches of complexity as emerging from transitions between an order and a disorder phase (Paperin *et al.* 2011).

In short, at the level of the brain, this exploration of the state space would correspond to a curious behavior of novelty exploration and openness to surprise, seeming to go against the FEP. However, K Friston and J Carhart Harris consider that this behavior, called "epistemic research" or "epistemic foraging", is induced by a learning objective, i.e., this behavior is allowed by the *a priori* that there is something to learn, a given expected uncertainty that must be reduced. Reducing this last level of uncertainty, and therefore learning, means choosing a policy that also maximizes the ability to predict through model selection (BMS and BMR) that makes the results less surprising (Friston *et al.* 2017; Carhart-Harris & Friston 2019). In this sense, a higher level of FEP, in the longer term, would generate exploratory behavior in the short term and be responsible for the experience of insight and intuitive understanding, whether during a psychedelic experience or not. Thus, in this perspective, the biological and human characteristics of curiosity, intuition and meaningful insight, essential to creativity, are always justified by the FEP.

The info-computational and cognitivist vision of life considers that any biological system is a computational process guided by the FEP, a principle of optimality oriented to preserving priors, homeostasis, and organization maintenance. If the entropic disorganization of living systems, by psychedelics in the case of the brain, allows the production of novelty, such a phenomenon is allowed only by a higher level following the FEP optimization.

However, we will see that the notions of information and optimization have shortcomings in life sciences. The FEP is based on assumptions (including the existence of a predefined phase space) that we think needs to be revised to understand living beings.

2. Criticisms and Limits of the Free Energy Principle Applied to the Living

2.1. Brief Philosophical Criticism of the Notion of Information

Information is at the core of many current technological developments, and it is tempting to go beyond this dimension and to think that the world is made of information and computations. Galileo followed a similar path when he stated that,

> Philosophy [i.e., natural philosophy] is written in this grand book — I mean the Universe — which stands continually open to our gaze, but it cannot be understood unless one first learns to comprehend the language and interpret the characters in which it is written. It is written in the language of mathematics [...] (Galilei 1623).

Instead, we think that it is necessary to distinguish our understanding from the things we aim to understand, that is, to distinguish physics (and biology) from metaphysics. The question we are addressing in this paper is theoretical: we seek to understand the world with conceptual tools, not to find its ultimate nature. From this perspective, the notions of information and computations are formal tools and concepts, not natural essences.

The info-computational paradigm is essentially based on the notion of Shannon information: in a given space of possibility, the possible signals to be transmitted, the amount of information, i.e., the informational richness, corresponds to the inverse of the probabilities of occurrence of a signal, in this sense, the rare is more informative than the frequent (Lesne 2014).

Boltzmann's entropy inspired this vision. However, Boltzmann's entropy has a coefficient – Boltzmann constant – that refers to a specific physical phenomenon requiring a physicomathematical interpretation, including units (Castiglione *et al.* 2008). Confusing information and entropy means forgetting this physical dimensionality. This misuse entails that information would be everywhere since entropy is produced wherever there is irreversible energy transformation. Brillouin's use is more interesting as it addresses information the experimenter can get from a system. He defines information as negentropy, that is, with the opposite sign w.r. to Shannon's information, and he argues that any measure that produces information requires a transformation of energy, therefore increasing entropy (Brillouin 1956). Information is physical in the sense that it requires physical transformations, but physics is not information; we do not think information is intrinsic to matter or that it has become a robust, fundamental concept of physics (Longo 2020). Notably, information is not associated with proper conservation principles.

The invariants of action constituted by a cognitive system, the belief structures from the perspective of the

FEP are the foundation of the notion of information, which creates a progressive detachment from the materiality of the phenomena that deepens with language, symbols, and even more by writing. Thus, cognition creates information from the contextual meaning, not vice versa. We distinguish information as the elaboration or transmission of signs and information as the production of meaning in active friction with reality.

In the informational approach, the production of meaning is the production of information. However, this approach eliminates intelligibility in favor of formal normativity by sets of instructions that would govern the living or by local optimization. This approach neglects the importance of interpretation and eliminates the biological singularity that comes from the historical formation of meaning by confusing salience and pregnancy. Salience has no meaningful depth; it is only a flat correlation, a regularity detection. It corresponds to what constitutes automatic learning algorithms, whose interpolative power finds regularities even in pure randomness (Calude & Longo 2017). By contrast, pregnancy already possesses elements of meaning, proto-semiotics embedded in the emotions and body (Sarti, Citti & Piotrowski 2022; Wildgen & Brandt 2010). In biology, pregnancy typically ultimately affects survival and the ability to reproduce.

Thus, in our view, constructing a hierarchy of meaning is not reducible to a formal question, to the results of computations constituting saliences in an optimized way. On the contrary, it is constituted by a practice of what is pregnant for the organism that acts for a purpose; it forces hierarchies of meaning on this basis. The brain is then a system of meaning production rather than information processing (Longo *et al.* 2012).

2.2. Criticism of the Principle of Optimality

The FEP can be understood as a physics of coupled systems (Ramstead 2023) and is based on two main physicalist assumptions, namely the optimization principle, grounded on the *a priori* of a pre-given phase space. There are relevant general criticisms that several authors have addressed (Longo, Montévil & Kauffman 2012; Montévil *et al.* 2016a; Sarti, Citti & Piotrowski 2019) included in (Colombo & Wright 2021; Guénin-Carlut 2023; Nave 2025). Here are the main points.

The principles of optimization presuppose the existence of an optimum locally or globally, in which case it serves as an attractor in the phase space and is determined *a priori*. This kind of reasoning is ubiquitous in physics to determine and predict dynamics. Even some dissipative systems far from equilibrium (e.g., clouds, hurricanes, or flames) are considered necessary and optimal geodesics in their phase space. Their forms are generic and not the result of a creative process, just like the configurations produced by algorithmic optimality methods. Accordingly, they can be generated de novo in practice. The phase space is predetermined, and all the possibilities are already there.

Thus, the main dynamics taking place are perpetuated and strengthened. The changes in configurations and the appearance of novelties take place only as a search for an optimum. This approach neglects the production of novelty in a strong sense, that is, involving a change of what is possible (Longo, Montévil & Kauffman 2012; Sarti, Citti & Piotrowski 2019; Montévil 2019). Similarly, assembly theory aims to understand how novelty can appear as a combination of the existing objects (Sharma *et al.* 2023). Thus, there is an opposition between creativity and mathematical optimality. This creativity, at the origin of the various survival strategies of an organism, does not pre-exist; there is no optimal way to discover it. Conversely, if we consider that the living produces new possibilities, optimization can have a meaning but is limited in its scope. When there are enough established and stable constraints to create a space with regular consequences, optimality can appear as an adjustment, primarily quantitative.

Let us unpack this concept of change of the possibility space. The way to model a change in physics is primarily a change of position in a predefined space, the state space. Some space changes exist, of course, in physics, from statistical mechanics fluctuations in the number of particles in the grand canonical ensemble to Fock's spaces in quantum mechanics; however, the condition for their use in modeling is that the new dimensions are theoretically identical to the old ones so that a single mathematical description can subsume them. By contrast, one of us has argued that biology requires addressing changes in possibilities that correspond more technically to situations whose organizational outcomes are not generic consequences of the causes established initially (Montévil 2019). In practice, it follows that what is possible cannot be prestated – even though we can, of course, prestate some possibilities.

This theoretical situation is the hallmark of historicity. The practical consequence is that the way to describe and manipulate theoretical living organisms is markedly distinct from physics, as exemplified by the names of systematics, which are defined by their historical origin and not by invariants of the causal determination (Montévil 2019). Without a proper account of this practical and theoretical consideration, there is a gap between the theoretical description and the empirical object of study.

The FEP also requires explicitly an assumption of ergodicity (Friston 2013), which has been criticized for biology (Longo, Montévil & Kauffman 2012). Ergodicity roughly means that the system will travel the possible states in a uniform way; and it is required to connect the microscopic and the macroscopic levels of description by means of entropy. The question of ergodicity is another way to look at the problem of predefined possibility spaces because ergodicity breaking corresponds to change in macroscopic possibilities.

On the other hand, within the FEP formalism, the result of a phenomenon depends on the path; one says pathway-dependent, a common approach in physics. Physics aims precisely to study what is generic and does not depend on context and history. In the case of pathway dependence, the past is integrated into the present, but only what has visible consequences on the path is taken into account. This approach does not retain what does not leave a visible trace in the final result; therefore, optimization levels down historicity. However, we understand biology only if we know evolutionary history, the past can re-emerge later in a contingent way and participate in generating new configurations (see Section 3).

Finally, the FEP needs to be revised in its relation to teleology. Indeed, variational principles, such as the principle of least action, can be interpreted teleologically, and this point is widely discussed (Glick 2023). In physics, there are counter-arguments to this teleological interpretation, but these counter-arguments are not relevant to the FEP. In physics, this principle can be seen as emerging; it is fundamental for FEP advocates. In physics, it corresponds to the stationarity of the action, so an extremum without specifying which, while it is a minimization for the FEP. Finally, the least action principle is about a trajectory, while the FEP explicitly sets the distal goal of a minimum of surprise. We do not think this assumption is necessary to biology, as we will see in Section 3.

Teleology raises different issues in physics and biology, and completely rejecting it in the latter is unnecessary. However, the teleology of the FEP considers a general purpose given by the FEP, which is problematic and constitutes, in our view, a conceptual regression concerning the historicity of the living coming from the theory of evolution. If there must be a biological teleology, it is very relevant to consider that living beings give themselves their own ends and that the latter can change over time. This point is precisely the proposal made by philosophers working with the closure of constraints: the organization can be interpreted as teleological because it self-determines through the circularity of the closure (Mossio & Bich 2017). The norms are then individual norms, which means that they can change. Moreover, the way closure changes also becomes historical and is not subsumed by an optimization principle.

2.3. The Cost of Optimality

According to the FEP, any living system is a nesting of Markov blanket, where each blanket defines a statistical partitioning between internal and external states. The internal generative model seeks to represent the external environment best in order to optimize its predictions and reduce the gap between what is perceived and what is expected. Thus, what is selected and observed preferentially is driven toward what can best validate the model's evidence and reduce its uncertainty. This situation amounts to an exploitative research behavior (Friston *et al.* 2017) consisting of being attracted only by what goes in the direction of the priors and denying or not paying attention to what is too distant from the expected, the things we do not know that we do not know. This optimization leads the sensory input to be similar to the output, thus the border between what is internal and external is transformed gradually in the impermeable border between the expected and the unexpected, i.e. the entropic alterity.

Moreover, more concrete actions on the world to reduce the uncertainty of the model also tend to reduce the unexpected and thus the possibility of learning genuinely new things. As a result, the priors are becoming stronger and less tolerant of uncertainty. In other words, the system becomes hyper-selective and only accepts what fits into the model and tends to stiffen it. This phenomenon is adequate to explain certain behaviors such as denial or confirmation bias, or even certain pathologies (depression, anxiety) where the world model takes over perceptions by conditioning them strongly (Badcock *et al.* 2017; McGovern *et al.* 2022).

Thus, the FEP alone necessarily induces self-referential confinement; Carhart Harris speaks of "conservation bias on adaptation" (Carhart-Harris 2018). This confinement can be compensated by a curious and exploratory behavior, requiring a certain acceptance of uncertainty, or by taking psychedelics, the two joining since psychedelics seem to encourage the exploratory behavior of the brain. According to REBUS theory, the increase in cerebral entropy by psychedelics "seems to breach the principle of free energy temporarily" (Carhart-Harris & Friston 2019), which appears beneficial to mental health and creativity (Mason *et al.* 2021). However, this breach is only apparent according to them. The authors evoke a new, higher level of application of the principle of free energy at the level of the models themselves (Bayesian Model Reduction and Bayesian Model Selection).

Thus, optimizing a higher level could explain the violation of the FEP at a certain level. The lower level, when it does not tend towards the optimum, would have an exploratory role because of the higher level. The latter would exercise the exploitative role necessary to speak of minimizing free energy and being causally responsible for curiosity. This induction of the local violation of the FEP would lead to new intuitive understandings. The upper layer would be responsible for this harmonious "recalibration" of beliefs, thanks to its operating FEP. Thus, the famous balance between exploration and exploitation, understanding and precision, or generalization and specification should be found in the interaction between two optimization layers.

However, the exploitation expected by the upper level induces and conditions the exploration of the lower level. The exploration is then remotely guided by the projection of what is helpful to discover and learn, which goes toward reducing uncertainty. This long-term orientation toward the optimal limits exploration, curiosity and will necessarily lead the system to shut itself from the unexpected.

Thus, a higher level of FEP does not, or only temporarily, counteract optimization excesses at the lower level unless it has an even higher level under the FEP and so on to infinity. These upper layers would be devoid of *a priori* in the form of belief except the intrinsic *a priori* of the FEP: a predetermined and, therefore, closed phase space and the pre-existence of optima imposing a finality.

Finally, in the FEP theory, exploration is an emerging phenomenon caused by its future optimality assumed by a higher level of FEP. Exploration is not a principle. There is no gratuitous curiosity; inferences act as motion-generating attractors, and the default state (without attractor) is immobility and conservation. Adaptation manifested as exploratory curiosity and learning, responds to a problem or a threat to survival, whether present or projected in the future as *a priori*. Necessity is the driving force of a transient contingency, just as invariance is the driving force of movement.

In Section 3, we will assume that exploration is a constituent of the default state of biological organizations and is revealed by suspending higher-level organizational constraints; therefore, the opening of a level does not require optimization at another level, the opening is constitutive. We propose to move from a computational Bayesian model to a more parsimonious theory of specific objects and constraints where we do not assume a general optimization principle.

3. For a New Biological Theory: Biological Organizations between Opposite Entropic Tendencies

In the continuity of Darwin's first principle, reproduction with variation, we elaborate a biology of the activity, motility, and changes in possibility that constitute the historicity of the living. We contribute to an alternative to the conservation and optimization principles inspired by the theories of inert objects, whose first assumption is the *a priori* definition of the phase space. Physics explains change on the basis of invariance; in biology, change is ubiquitous and we need to explain historicized invariances; see (Montévil & Mossio 2020). The relationships that constrain and enable the organization and evolution of life constitutes our theoretical starting point. In this section, we will analyze two aspects of entropy and use this analysis and the subsequent concepts to discuss biological situations where, we contend, entropy is not defined as a function in general.

3.1. Two Opposite but Complementary Entropic Trends

1) Physicists generally speak of entropy increase when there is a dispersion of energy. Entropy increase corresponds, at the local level, to an increase in the number of equivalent microscopic states, thus a form of randomness, and, at the global level, to homogenization. For example, when particles of a gas are concentrated at a given location, it tends towards a uniform distribution. The latter has a simpler macroscopic description than the former – we do not need to specify the location describing the heterogenous distribution –, and it corresponds to far more microscopic configurations, because all particles have the same chance to be anywhere. Entropy increase describes processes directed in a single direction, the most generic one, which amounts to tending towards a form of stability and predictability. In the structure of thermodynamics and statistical mechanics, entropy is used to specify the final state of a system tending to equilibrium, as the one with maximum entropy satisfying the constraints.

In a statistical physics system, in a sense, entropy and energy functions compete because they have opposite signs and thus effects. When the temperature is high, entropy, in the form of random agitation of particles, dominates, for example, in a gas. When the temperature is low, the energetic constraint dominates leading for example to a crystal. However, in both cases the above discussion still applies and entropy remains structuring. The system tends towards the most generic macroscopic state given the internal constraints, energy in particular, and the external constraints as the boundary conditions.

Then, local randomness gives stability at a larger scale; the homogenization of the local variation then justifies stability. The two trends described by entropy increase go together but they also have opposite meaning – increase of the predictability at the global level, and decrease at the local level, for example. While variation is generally associated with disorder, homogenization and stability are generally associated with "order." There is, therefore, a form of "ordering" described by entropy increase at the global level, as Schrödinger already envisaged in his notion of "order-from-disorder" (Schrödinger 1944), where order simply means macroscopic regularities.

2) We can find this double entropic trend in the FEP and its application to the Bayesian brain. Indeed, the FEP assumes that any living system, particularly the brain, follows a Bayesian generative model of world representation that evolves by maximizing its evidence and the validity of its representations, which amounts to minimizing its relative entropy. Here, the local entropic trend is described by the relative entropy, also called uncertainty, surprise, or prediction error. The FEP is a variational principle, like the least action principle and the maximum entropy principle, but applies to Bayesian mechanics as "physics of and by beliefs" (Ramstead *et al.* 2023). The FEP states that the generative belief model always evolves in the direction of the most probable, ultimately tending to a stationary state of maximum entropy that can be interpreted as the global entropic trend of the system under the constraint of the *a priori* of the model and external states. Thus, according to the FEP, living systems are teleologically oriented models, following a physics of beliefs toward their stability in their environments driven by the global entropic trend. They oppose the dissipative local entropic trend that is also a source of "information" on their environment. This modeling is organized as nested Markov blankets where the global level operates to optimize, notably simplifying and reducing the local level according to the regime of constraints made of *a priori* beliefs.

In the previous section, we pointed out some limitations of this vision. Firstly, the phase space is already predetermined, which means that the exploration of new possibilities is limited. Secondly, the process of optimization overrides historicity, which means that the context and history of the object are not properly taken into account. Finally, an exploration that is induced and determined by a higher layer of optimization does not compensate for the self-referential confinement. We propose an alternative perspective that shares some similarities with the FEP approach but has fundamental differences.

3) We extend the definition of these two entropic trends by their level of description. In physics, the global is derived mathematically from the local by a state or path integral, with the assumption of ergodicity. In biology, we introduce a concept of global entropic trend that would not derive mathematically from the local entropic trend, because we do not assume ergodicity and a fixed possibility space. Let us insist that in biology we do not assume that the global entropic trend corresponds to a state function; however, we import, in biology, part of the conceptual role played by entropy

in physics, specifically in the tension between the local and global aspects of entropy. The aim of this extension is to work out the articulation between local and global, variations and stability.

We will present briefly our perspective on biological organization to go further. The idea is not to give a fixed definition of biological entities, which are always transient and historical, but to think about the processes that give rise to their relative stability (Soto, Longo, Miquel *et al.* 2016). This involves considering their historicity and defining the organization of constraints within the living, according to a "closure of constraints" constituted historically and contextually (Montévil *et al.* 2016a; Mossio, Montévil & Longo 2016). This concept of closure differs from the concepts of convection cells or catalytic cycles. It also further specifies the autopoiesis of Maturana and Varela, since the latter does not specify the theoretical entities that are the subject of this self-production. The specificity of the closure of constraint is to define constraint dependency where the recursion in the constraint chain "folds and establishes a mutual dependence." Constraints maintain and compensate collectively for their dissipation through constrained processes (Montévil & Mossio 2015), generating and regenerating their interdependencies and leading to processes that would not occur without constraints. It is this mutual maintenance that enable biological organizations to last over time and diversify, by contrast with physical self-organization that is a spontenous self-organization of flows. The relevance of a given constraint remains limited in time and contingent; organization can change over time. Closure of constraint implements the coordination of interactions and relationships within living beings and with their environments. However, their environments do not fully determine them, and they resist them in a certain way. Thus, the network of constraints can collectively determine itself, that is, it self-maintains by self-constraint (Moreno & Mossio 2015). This framework also provides a concept of biological function (Mossio, Saborido & Moreno 2009).

Starting from this approach of organization, we define the global entropic trend as the "attractive" tendency of closure of constraints to shut itself from destabilizing influences and achieve a stationary equilibrium, i.e., a state less and less likely to evolve, neither under the influence of a (relatively small) external disturbance nor internal local fluctuations. This change is, therefore, a trend towards the most likely state related to the priors and the context. The constraint regime formed up to now is simplified by maintaining its link in the context and then, at the limit, maintaining itself identically. In this slow and gradual evolution towards stability, local entropic trend, fluctuations that can affect the closure of constraint, is minimized. The evolution of the closure of constraint following this tendency tends to behave as if it followed a trajectory in a state space co-determined by the relationship between its historicity and the context. Thus, the more a closure of constraints closes to the variation, the more the phase space and probabilities are definable. No entropy function is defined in the general associated case, but in this limit case, we may consider an associated function that could be that of the FEP. This limit case is also the reason we introduce the general terms of global and local entropic trends in a situation where there is no general corresponding state function.

Closure justifies at least in part the relative stability of the constraints involved, and as such, it participates in global entropic trend. However, once organized in closure, the global entropic trend also reinforces stability by self-simplifying (Umerez & Mossio 2013). This concept of global entropic trend can be used to talk about the invariances observable in the living, the tendency to maintain, reproduce, repeat, and homogenize. However, it is a trend and not a state reached in the living because any closure of constraint that would be too shut from the milieu becomes fragile, loses plasticity, and risks destroying itself abruptly with no possibility of resilience, which is encountered in second-order disruptions (Montévil 2022) (see Section 3.3). Let us also insist that this tendency to shut oneself differs from the thermodynamic sense of a closed system, an organization being always open from the latter point of view. It corresponds to the absence of change of organization by friction with the environment. Thus, the tick described by Von Uexküll reduces his world to a minimal number of relevant aspects and has mostly automatic responses to these aspects (Jakob von Uexküll 1965).

Let us emphasize that the global entropic trend should be analyzed at the level of the closure of constraints and that the latter always has a global dimension. It constitutes a higher level of organization compared to the constraints that constitute it, considered as local, multiple, diverse, and can be affected by local entropic trends.

In biology, we posit that local entropic trends corresponds to any variation affecting a closure of constraint that is not part of a pre-established possibility described by this closure and compatible with it. In this sense, the local entropic trend corresponds to unpredictable variations with respect to the knowledge of the initial situation (Longo, Montévil & Kauffman 2012). The local entropic trend is a source of closure of constraint's changes and manifests itself at different scales, such as "infidelities of the milieu" (Canguilhem 2013); these can be external, i.e., environmental or internal: DNA mutations, rare configurations and interactions of proteins (the stereo-specificity of the macromolecules being only partial), the random and asymmetric distribution of cellular components between two daughter cells, allelic rearrangements during meiosis, the recombination of ancestral phenotypes, new neural connections, neural spontaneous activity…

By contrast, already functionalized randomness is not a genuine contribution to the local entropic trend from the biological perspective. For example, stochastic resonance is a situation where adding randomness contributes to functionality by amplifying a signal by adding noise. Similarly, most molecule movements in a cell directly result from diffusion. In these examples, randomness does not destabilize the initial organization, it contributes to it.

The global entropic trend corresponds to maintaining the system and repeating at least some of its dynamics, allowing living organisms to maintain homeostasis and regulate their internal functions. On the other hand, local entropic trend leads to the divergence of the system and its disorganization. Therefore, we argue that living organisms exist in a tension between local and global entropic trends.

3.2. Anti-entropy as the Tension Between Global and Local Entropic Trends

In the living, global and local entropic trends are in tension between homogenization by the global, that is to say, by the organization at the larger scale, and heterogenization by the local, where one canalizes the other. For example, the ecosystem can exert a relatively stable environmental selection pressure relative to the life of an organism; at the same time, each new organism brings its contribution to variation, both in relation to other organisms of the same species and in relation to its viability in its ecosystem. This negative selective pressure is globally homogenizing; it stabilizes the local entropic trend in specific functional configurations. Thus, there is not a single optimal configuration, an organism stereotype, to select but a set with a common characteristic to be sufficiently viable in their ecosystemic context. This process of openness to variation and homogenization by excluding the incompatible is found at every scale of the living. This process creates the diversity of individualities and types of individuality (organism, cell, ecosystem, species), appearing homogeneous on a global scale but having an internal diversity that can manifest itself in a different, unusual context.

However, this tension is not in equilibrium or search of equilibrium; it is even less 'optimal' or 'perfect.' On the contrary, there is a discrepancy, a non-identity of oppositions, and a certain "relaxation" of constraints, which generates a continuously reorganized dynamic. This consists in changes of symmetry in cascade linking different levels of organization and allowing these levels of organization to exist. In our view, organization is more than near-critical, as describe in (Safron *et al.* 2022; Parerin *et al.* 2011), where organization emerges in "edge of chaos" inter-regimes balancing between disordered and ordered dynamics, in a pre-defined ''phase space" containing the trajectories of the dynamic. Similarly, we argue that since the space of possibilities is constantly transforming (which is incompatible with formal optimality), we say that organizations are then in a state of extended critical transition (Longo & Montévil 2011).

In this context, we consider that the production of anti-entropy, that is to say the appearance of new functional possibilities, takes place between two opposite tendencies: to approach global entropic stability, a trend to "optimality," and to move away from it by local variation. It maintains its imbalance, a sort of back and forth between these two crucial but destructive tendencies in their limit case (see Section 3.3). Anti-entropy production would then be in the tension, never resolved, between local and global entropic trends.

This tension is found between the tendency to conservation, identical reproduction, and repetition by the stability of the global and the "open-ended" evolution, the divergence by the variability of the local. Each living system has a certain degree of stability for its maintenance. However, it also requires variation to maintain its internal diversity, which tends to disappear by homogenization (see Section 3.3). This idea is found in the concept of proliferation with variation, driving evolution and ontogenesis (Soto, Longo, Miquel *et al.* 2016). Even if the part of essentially similar reproduction is greater than the part that varies (some mutations in the case of meiosis, asymmetric distribution of some constituents in the case of mitosis), the combination of the two is necessary for life. In short, maximum homogenization and variation are like two entropic "attractors" but at different levels, respectively global and local.

From this perspective, the production of anti-entropy, as the appearance of functional novelty by integrating variation into an organization, still requires the local entropic trend in the form of variation. Consider, for example, one of the major evolutionary transitions: the formation of eukaryotic cells by the symbiosis of bacteria and archaea, particularly the formation of mitochondria (Martin, Garg & Zimorski 2015). The invagination of a bacterium in an archaeon, leading to the appearance of mitochondria, was a large entropic disturbance for the host archaeon or even for both. Then, by co-evolution, their relationship became symbiotic, a new viability situation among many failures of this evolutionary 'accident.' This case exemplifies diversity production through the entropic encounter of distinct evolutionary paths (Longo, Montévil & Kauffman 2012).

In general, of course, no variation is directly anti-entropic. Its effect in time and space on the existing biological organization causes successive symmetry changes, i.e., processes of disorganization requiring reorganizations. Think again of cell division: the entropic component of proteome distribution, of partial DNA repair, contributes to the anti-entropic production of the new organization, generating diversity. If the reorganizations make it possible to maintain this new organization, the entropic variation was then transformed into a functional anti-entropic novelty. It is then at the origin of evolutionary diversity at different levels of biological organization. However, it also contributes to tissue differentiation during embryogenesis – through strong sensitivity to contour conditions (pressures, biochemical flows, etc.). Thus, it is as if anti-entropy is "nourished" by the local entropic trend.

The notions of "flow of variety" and "stasis" in Nietzsche's philosophy could be associated, respectively, with that of local and global entropic trends. According to him, "the flow destroys inherently the means implemented by life to protect itself" (stasis). However, this flow feeds life by "incorporation," allowing it both to maintain itself and to evolve its stasis. The variability of the flow and the stability of the stasis are then in "tragic tension" (Stiegler 2021).

Entropic variability, which generates "defects" compared to the norm of a living system, is necessary for its evolution and, therefore, for evolution in all its forms, including learning.

Note that the new organization is not necessarily more complex than the one from which it comes; there is no teleology towards ever more complexity. However, complexity may allow for the invention of new ecosystemic niches, in which case it is more likely to survive (Cazzolla Gatti *et al.* 2018).

3.3. From the FEP to the Anti-Entropic Vision of Life

Let us now emphasize the fundamental differences between our proposal and the FEP.

The Toxicity of the Global Entropic Trend

Although the global entropic trend is an essential component of life, by limiting the local entropic trend, it also can have negative consequences. This trend does not end at the maintenance of the closure of constraints; it continues to strengthen by self-simplification of a closure. This tendency, when extreme, can lead to harmful consequences for living beings in two linked and mutually reinforcing ways: the reduction of the richness of historicity and the closure to the contingency of the real, understood as what resists, especially to representations. The real is nevertheless a source of historicity and, therefore, of anti-entropy production.

Indeed, reducing the entropy of the past, of the belief model itself in the FEP language, amounts to erasing its details, i.e., its internal diversity from history, by a semantic oversimplification. The excessive loss of the memory traces of the contextual elements in which the closures were built reduces the historical richness to a single trajectory and generic behavior devoid of tensions. This consolidates the oldest and most general traces by eliminating redundancies. The object becomes more and more predefined by an optimized pattern, able to predict in a fixed milieu and only open to the proximal, restricted future.

This trend results in intense selective pressure on the integration of the local variations. This closedness to the present diversity of reality reduces the integration of the local variations and the depth of integration. Thus, the global entropic trend tends to make the closure of constraint necessary, structural, rigid, automatic and ahistorical. It tends to abstract itself and become impermeable to the context by building a space of possibilities closed and in contraction. This phenomenon leads to a loss of function by loss of the frictional relationship with the contingent complexity of the context, leading, in short, to self-referential confinement.

In other words, we move from metastability to stability. The excess of constraint by the higher scale and the submission of the lower scale leads to eliminating its degrees of freedom supported by the various redundancies. There is a leveling down of levels into a synchronic unit exceedingly coherent with itself but detached from the real context and its improbable contingency. Diversity is reduced and canalized into hyper-specialization. These phenomena lead to a loss of the plastic resilience of the organization, i.e., its ability to produce anti-entropy by integrating local entropic variation. The result is a reduction in the space of possibilities that can go as far as second-order disruptions, that is, the loss or impairment of the ability to produce functional novelty (Montévil 2022).

The FEP, when considered the only fundamental law, entails the strengthening of a model by permanent research of validation of the model's proofs, leading to self-referential confinement. Concerning cognition, this is manifested by a rigidification of thoughts, which is found in several psychopathologies (e.g., physiological aging, end-of-life depression, Alzheimer's) where the activity of DMN is strengthened (Cieri & Esposito 2018). Additional levels of FEP do not eliminate these problems (see Section 2.3).

The rigidification and the reinforcement of the closure of constraint are limited by the local entropic trend, which tends to destabilize them by bringing variation and making them evolve. This idea is common to FEP and our approach. However, for us, the local entropic trend is not only external or "accidental," i.e., due to the organization's instability. Moreover, the living not only repels variation but also maintains an open relationship with the "unprestatable" (Longo, Montévil & Kauffman 2012). Local entropic trend cannot be modeled by injecting an amount of randomness into a model.

Organization and Local Entropic Trend

Local entropic trend not only includes disturbances coming from outside but also those of internal origin, i.e. brought by historicity. In the latter case, it corresponds to traces of history not fully functionalized to the system and can be reinterpreted into new functionalities according to the context. The traces of the past resist normalization; they constitute a form of internal diversity, of alterity within the system itself. Memory then constitutes a reserve of deviation and not only a united block conditioning the future towards ever more optimality. We can refer to the exaptation of vestigial structures as an example. A past structure reappears but not wholly; it is reinterpreted according to context and may result in a new organizational function (Rayner, Sturiale & Bailey 2022). Let us emphasize that its potentialities coming from traces of the past are not like hidden possibilities whose properties are actualized. Their reinterpretation in the present gives them a new biological meaning. Thus, what matters to the organization at a given moment only partially defines what it is for the next moment. The new organization cannot be formalized from the previous one because of this incompleteness resulting from historicity.

On the other hand, we argue that living beings maintain a fundamental openness to the local entropic trend that comes from the principle of variation (Montévil *et al.* 2016a). This openness also appears in what has been called propulsive constraints (Miquel & Hwang 2016; Montévil & Mossio 2015; Montévil & Mossio 2020). Their role is to actively open the system to variation, which goes against the FEP. For example, we can cite all the constrained processes (more or less dependent on the context) bringing novelty when generating a new organism: in bacteria, the modulations of genetic mutations according to the context, exchanges of genetic material; and in protozoa: crossing over, random phenomena during sexual reproduction.

Proponents of FEP could argue that evolution would have optimized its propulsive constraints. There is probably some optimization, but it occurs after the appearance of a novelty, including a second-order evolution novelty (Tenaillon *et al.* 2001), and does not explain its emergence. It requires a first opening to alteration, a relaxation of constraints not guided by a superior optimality.

These considerations lead us to discuss what activity and passivity mean in the living.

New Perspective on Activity and Passivity

From the perspective of the FEP, passivity is the reception of sensory input from outside; activity is active inference in two forms: transforming the environment or transforming the representation of the environment to make it less surprising in the future.

On the other hand, we propose that passivity corresponds to the mechanical, predictable functioning already included in the pre-established dynamics by the constraints of the organization. Thus, the "active" inference, made by the automatic projection and transformation of expectations according to the FEP, is then also passivity in the sense that its dynamic is pre-established towards the minimization of variation.

To explain this, let us return to FEP: it is not specific to the living and is considered as a "physics of beliefs" (Ramstead *et al.* 2023). Just as in physics, objects are passive with respect to the laws governing them; biological organization is passive with respect to the law described by the FEP. When the organization complies with the FEP, that is to say, when the closures of constraints constituting it are simplified without functional innovation, the evolution takes place "mechanically" within a space of possibilities predefined, thus in a kind of passivity. The case of the physicist's "active matter", as described in (Chvykov *et al.* 2021), corresponds to a statistical mechanics where the particles are out of equilibrium, but it also corresponds to passivity in our perspective because the particle follow fixed rules.

On the contrary, there is activity when there is a change in the organization in a strong sense. This implies an active opening to variation followed by a reorganization on several scales. The organization is actively involved in its transformation, outside a space of predefined possibilities, by integrating the local entropic variation that is not governed by a "law" (Tahar 2023).

3.4. The Question of the Default State of the Living

The question of the biological default state has been put forward by (Sonnenschein & Soto 1999), notably by analogy with the principle of inertia, a state at the basis of classical physics. Inertia is never exactly observable, but it structures the theory. The theoretical strategy is analogous for the default state in biology.

In the theoretical perspectives inspired by physics, the current paradigm of the default state of living systems is self-preservation (Bourgine & Stewart 2004), similarly a common assumption in biology is that the default state of cells is quiescence (Montévil *et al.* 2016b). Change appears as a means for maintenance; this hypothesis goes hand in hand with the search for balance, stability, and optimality, imposing a limit to the changes. For example, at the cellular level of multicellular organisms, this default state would manifest as quiescence, an inactive cell at rest waiting for a triggering stimulus.

Soto and Sonnenschein initiated a reversal of perspective by assuming that the default state of cells is proliferation and motility and not quiescence. It follows that there is no need for stimulation for cells to display this default state. Instead, quiescence requires an explanation in the form of a cause (Soto, Longo, Montévil *et al.* 2016).

In our approach, we suggest moving from the primacy of the teleological principle of entropy minimization to the notion of anti-entropy as a tension between local and global entropic trends. According to this notion, living beings are not intended to reduce entropy to the maximum but to maintain a degree of openness to feed on it, that is, to functionalize it and transform themselves. Living beings are not just fighting against the local entropic trend but, instead, grow from it.

We go from a default state of least action, passivity, to a default state of activity where the variation is not triggered in response to a disturbance from the outside but is intrinsically present, canalized, and more or less maintained. It is a state of exploration outside a predefined phase space without reward and not constrained by a superior organization. This exploration, requiring an intrinsic openness to variation, takes place not only through genetic variability but also at different levels of life; it appears as motility, mobility, or curiosity not motivated by a goal.

This exploratory impulse, most of the time repressed and constrained, does not stem from a superior commitment to optimization. Considering this as a default state has consequences on causality: if we assume that the activity is by default, then if it is not observed, it means that it is constrained, and we have to make these constraints explicit.

3.5. The Relationship to the Space of Possibilities

The principle of variation and the interplay between local and global entropic trends that we have just presented underlies a space of possibilities in continuous transformation by default and whose dynamics are not pre-definable (Montévil *et al.* 2016a). This is a fundamental difference between the living and inert, and also between biological and algorithmic creativity.

In algorithmic combinatorics, the elements are defined, distinct, and preexisting in a synchronic and complete co-presence. They are then combined according to pre-established rules with more or less randomness, which is itself predefined. The resulting "creativity" amounts to actualizing a part of a space of possibilities that is already defined, typically on the basis of a gigantic amount of data.

In biological generativity, there can be a combinatorics of elements and randomness. However, this process is diachronic because the "elements" are, in reality, a tissue of relationship, a set of constraints possessing a singular, meaningful historicity. They then have the possibility of being destroyed, mixed, entangled by the local entropic trend, and then "rewoven", allowing them to cross the barriers of the probable and thus open the space of possibilities. Of course, this view is a major epistemological challenge, which is why assembly theory assumes generic properties for selection that remain hypothetical and independent of the nature of the novelties appearing.

Let us take the example of a mutation on a DNA nucleotide. There seems to be a defined combinatorics, so a determined phase space. However, this space is not sufficient for the theoretical determination of the object. Its functional consequences will depend on the neighborhood of this nucleotide (if it is in a gene, if there are several modes of reading the gene), the epigenetic structure (e.g., accessibility of the gene, localization in the nucleus), the cellular context (e.g., cell type, cell neighborhood) and the context of the organism and its environment. All these layers of organization and their space of possibilities have a certain degree of determinism necessary to maintain them. However, the entanglement between the different levels opens the space of possibilities to the indeterminate, contributing to its expansion.

Conclusions

The question of a theoretical framework to understand organisms is an open debate. In this debate, one of us has contributed to developing three theoretical principles: the principle of variation, the principle of organization, and the default state (Soto, Longo, Miquel *et al.* 2016). These principles are sufficiently robust to be foundational in biology; nevertheless, they are also starting points, and much remains to be elucidated, notably concerning how biological organizations change.

In that regard, a parallel effort has been accomplished on the notion of a Free Energy Principle (FEP) stemming from cognitive sciences and based on an informational perspective. The FEP is a framework that explains how a system and its "belief" model evolve through Bayesian updating. This updating is guided by an optimization principle that involves adjusting the statistics of the things to which they are coupled. This info-computational approach provides a view of self-organization where organisms are layers of abstract representation that generate probabilistic decisions. These representations are created empirically by detecting common patterns, followed by a succession of reduction operations, leading to model changes. The temporal evolution of a system is considered as a combination of a deterministic component and a noise component, which must be minimized. Noise is considered a source of novelty, as in the case of the entropic brain, where entropic annealing is used to explore the phase space and update new configurations.

From the general perspective of the theory of organisms, we have argued that the FEP is not acceptable as is. It assumes a pre-given possibility space, which is the condition of possibility of an optimization principle. In contrast, the principle of variation posits that biological possibilities change over time and rejects general optimization principles. Moreover, the general informational perspective of the FEP is problematic and has been heavily criticized by others. Specifically, the FEP leads to systems that would strengthen their models in the context of their coupling with their environment by minimizing surprise and uncertainty. As a result, it struggles to address putative beneficial situations where the brain entropy increases, like in the model of the entropic brain under psychedelics. For FEP proponents, the way out is to propose a schema of nested optimization levels so that increased entropy at one level would be part of optimization at another level. This way out is problematic, though, since this higher level would also require a phase space and regularities to optimize entropy. In any case, it remains incompatible with the notion of changing phase space and historicity as we define it.

Nevertheless, the perspective of the FEP can be analyzed by concepts that we introduce here, namely the distinction between *global and local entropic trends*. In physics, the local entropic trend would be the tendency of the system towards the microscopic fluctuations, while the global entropic trend would be the tendency of the system towards the macrostate that is directly or indirectly determined and stable by the second principle of thermodynamics, that is to say, the most generic state. By generalizing these concepts in the context of the theory of organisms, the local entropic trend corresponds to changes that are not yet functionalized, irrespective of whether they are of intrinsic or extrinsic origin. Global entropic trend corresponds to a tendency toward homogenization provided by a given organization and its coupling with its milieu. Then, the FEP considers only the global entropic trend in a specific informational setting, while the theory of organisms includes a principle of variation that, in the terms of this article, posits the universality and ubiquity of the local entropic trend.

In the theory of organisms, the global entropic trend partly corresponds to the closure of constraints, understood as stabilizing constraints and the corresponding processes. However, the notion of global entropic trend is more general. It opens the perspective of the tendency to simplify a closure, preserving and stabilizing its main functions and couplings with its milieu, possibly by taking inspiration from the FEP. Now, local entropic trend, of course, is related to the principle of variation. However, the principle of variation is about functional variation, while the local entropic trend is about variations that may be functionalized. As such, for example, we have emphasized the traces of the past that are not functional for a given organization, and that would be leveled down by the global entropic trend, but that may also enable new functions.

In a nutshell, the core message of our work, in line with previous discussions (Montévil & Mossio 2020; Longo & Montévil 2012), is that biological organizations are not, and do not tend to, organizational fixed points. Instead, they are between two opposite tendencies: the global entropic trend of homogenization and the local entropic trend of destabilization.

As a perspective, from the analysis of the two approaches mentioned, FEP and organicist, we can conjecture that there are two modes of biological evolution:

- Passive: exploring the space of possibilities already defined by a set of constraints (previous belief, inclination, habitus). This mode of development, governed by the FEP, advances cautiously and incrementally by capitalizing on what already exists.

- Active: that is, creative in the strong sense. It involves the change of the set of constraint by the local entropic variations and is able to change the space of the possibilities. It is actively promoted by the propulsive constraints.

The conjunction of these two modes of evolution creates relevant organizational changes, i.e *anti-entropic* in that it induces *a virtuous circle of viability by allowing both stability of the organization and openness as a possibility of new changes*.

What are the Consequences for Cognition?

According to the current consensus, the brain's default state (when the DMN is activated) is linked to the ego as a medium of identity. This autobiographical self, supported by all the memories (representations), guarantees the stability of the sense of identity despite the perceptive changes (Damasio 2000) by minimizing free energy (Carhart-Harris & Friston 2010). It is, therefore, a state of constraint and self-maintenance.

Our perspective leads us to consider a different default state for the brain or, more generally, the cognition of living beings. In contrast to the DMN, which focuses on self-preservation, it would be a state of exploration activity and curiosity free from its constraints in the form of belief. At the biological level, it would be a state where neurons activate themselves and make spontaneous connections in a contingent way, as it is the case for unconstrained cryptic 'spontaneous electrical low-frequency oscillations' SELFO (Hanson 2021), and at the psychological level, a state of creativity. This default state, constrained and therefore repressed in everyday life, can possibly be experienced, among others, through the experience of ego death reached during psychedelic experiences or deep meditation. In this case, psychedelics and meditative practice would not act as triggers of this state but as a relaxation of constraints on this state, which corresponds to Carhart-Harris's thesis and his hypothesis of the entropic brain.

The question arises of the re-organization of constraints, which are preserved, destroyed, or created; how do they change in a relevant way, i.e., therapeutic? According to the FEP, these are the ones that minimize uncertainty the most. According to our anti-entropy approach, on the contrary, constraints are reorganized in order to generate a greater capacity for openness to uncertainty. This openness can involve the removal of the most restrictive constraints (beliefs related to depression, for example) but, above all, a greater capacity for acceptance of the contingency of oneself and the world, like in stoicism. This capacity corresponds to an increased ability to generate anti-entropy from local entropic variations. At the psychological level, it can manifest as greater confidence in the becoming, which precisely does not rest on beliefs because it comes before the constitution of beliefs itself.

The lifting of blocking constraints can be learned through different techniques, for example, meditation (Ho, Nakamura & Swain 2020) potentially complementary to the use of psychedelics. In a sense, philosophy or even sciences are also methods and attitudes that require such an openness. All these techniques of relaxation of constraints and openness to contingency require double attention to the sensitivity and suspension of judgment, a fundamental gesture in philosophy also called "epoché" (Guilielmo & Mudry 2021). This voluntary and active work can be considered as a propulsive constraint, since it is a question of organizing its disorganization in an undirected way. Thus, perhaps one of the main lessons of these practices is to realize that seeking to minimize uncertainty is a locking belief, while it is liberating to accept it.

Acknowledgments

This work is supported by the Interdisciplinary program: Theoretical Foundations of Biology.

We are grateful to Giuseppe Longo and Anton Robert for their critical insights on a previous version of this manuscript.

References

Amsterdamski, S 1990 *La Querelle Du Déterminisme*. Paris: Gallimard.

Badcock, PB, Davey, CG, Whittle, S, Allen, NB, & Friston, KJ 2017 "The depressed brain: An evolutionary systems theory", *Trends in Cognitive Sciences*, vol. 21, no. 3, pp. 182–194. https://doi.org/10.1016/j.tics.2017.01.005.

Badcock, PB, Friston, KJ, Ramstead, MJD, Ploeger, A, & Hohwy, J 2019 "The hierarchically mechanistic mind: An evolutionary systems theory of the human brain, cognition, and behavior", *Cognitive, Affective & Behavioral Neuroscience*, vol. 19, no. 6, pp. 1319–1351. https://doi.org/10.3758/s13415-019-00721-3.

Bailly, F, & Longo, G 2009 "Biological organization and anti-entropy", *Journal of Biological Systems*, vol. 17, no. 1, pp. 63–96. https://doi.org/10.1142/S0218339009002715.

Bailly, F, & Longo, G 2011 ''Mathematics and the natural sciences: The physical singularity of life", *Advances in Computer Science and Engineering*, vol. 7. https://doi.org/10.1142/p774

Bourgine, P, & Stewart, J 2004 "Autopoiesis and cognition", *Artificial Life*, vol. 10, no. 3, pp. 327–345. https://doi.org/10.1162/1064546041255557

Brillouin, L 1956 *Science and Information Theory*. New York: Academic Press.

Buckner, RL, Andrews-Hanna, JR, & Schacter, DL 2008 "The brain's default network: Anatomy, function, and relevance to disease", *Annals of the New York Academy of Sciences*, vol. 1124, no. 1, pp. 1–38. https://doi.org/10.1196/annals.1440.011.

Calder, AE, & Hasler, G 2023 "Towards an understanding of psychedelic-induced neuroplasticity", *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, vol. 48, no. 1, pp. 104–12. https://doi.org/10.1038/s41386-022-01389-z.

Calude, CS, & Longo, G 2017 "The deluge of spurious correlations in big data", *Foundations of Science*, vol. 22, no. 3, pp. 595–612. https://doi.org/10.1007/s10699-016-9489-4.

Cameron, LP, Benetatos, J, Lewis, V, Bonniwell, EM, Jaster, AM, Moliner, R, Castrén, E, McCorvy, JD, Palner, M, & Aguilar-Valles, A 2023. "Beyond the 5-HT2A receptor: Classic and nonclassic targets in psychedelic drug action", *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, vol. 43, no. 45, pp. 7472–7482. https://doi.org/10.1523/JNEUROSCI.1384-23.2023.

Campbell, JO 2016 "Universal Darwinism as a process of Bayesian inference", *Frontiers in Systems Neuroscience*, vol. 10, article 49. https://doi.org/10.3389/fnsys.2016.00049.

Canguilhem, G 2013 *Le normal et le pathologique*. Paris: Presses Universitaires de France. https://www.cairn.info/le-normal-et-le-pathologique--9782130619505.htm.

Carhart-Harris, RL, & Friston, KJ 2010 "The default-mode, ego-functions and free-energy: A neurobiological account of Freudian ideas", *Brain: A Journal of Neurology*, vol. 133 (Part 4), pp. 1265–1283. https://doi.org/10.1093/brain/awq010.

Carhart-Harris, RL, & Friston, KJ 2019 "REBUS and the anarchic brain: Toward a unified model of the brain action of psychedelics", *Pharmacological Reviews*, vol. 71, no. 3, pp. 316–344. https://doi.org/10.1124/pr.118.017160.

Carhart-Harris, RL 2018 "The entropic brain – revisited", *Neuropharmacology, Psychedelics: New Doors, Altered Perceptions*, vol. 142 (November), pp. 167–178. https://doi.org/10.1016/j.neuropharm.2018.03.010.

Castiglione, P, Falcioni, M, Lesne, A, & Vulpiani, A 2008 *Chaos and Coarse Graining in Statistical Mechanics*. Cambridge: Cambridge University Press & Assessment https://assets.cambridge.org/97805218/95934/frontmatter/9780521895934_frontmatter.pdf

Cazzolla Gatti, R, Fath, B, Hordijk, W, Kauffman, S, & Ulanowicz, R 2018 "Niche emergence as an autocatalytic process in the evolution of ecosystems", *Journal of Theoretical Biology*, vol. 454 (October), pp. 110–117. https://doi.org/10.1016/j.jtbi.2018.05.038.

Chollat-Namy, M, & Longo, G 2022 "Entropy, negentropy and anti-entropy: Game of tensions to think the living", *ISTE OpenScience*. https://www.di.ens.fr/users/longo/files/ChollatLongo-entropies.pdf.

Chvykov, P, Berrueta, TA, Vardhan, A, Savoie, W, Samland, A, Murphey, TD, Wiesenfeld, K, Goldman, DI, & England, JL 2021 ''Low rattling: A predictive principle for self-organization in active collectives", *Science*, vol. 371, no. 6524, pp. 90–95. https://doi.org/10.1126/science.abc6182

Cieri, F, & Esposito, R 2018 "Neuroaging through the lens of the resting state networks", *BioMed Research International*, article 5080981. https://doi.org/10.1155/2018/5080981.

Colombo, M, & Wright, C, 2021 "First principles in the life sciences: The free-energy principle, organicism, and mechanism", *Synthese*, vol. 198, no. 14, pp. 3463–3488. https://doi.org/10.1007/s11229-018-01932-w.

Czégel, D, Giaffar, H, Tenenbaum, JB, & Szathmáry, E 2022 "Bayes and Darwin: How replicator populations implement Bayesian computations", *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, vol. 44, no. 4, article e2100255. https://doi.org/10.1002/bies.202100255.

Damasio, A 2000 *The Feeling of What Happens. Body and Emotion in the Making of Consciousness*. London: Vintage.

Dawood Hristova, JJ, & Pérez-Jover, V 2023 "Psychotherapy with psilocybin for depression: Systematic review", *Behavioral Sciences* (Basel, Switzerland), vol. 13, no. 4, article 297. https://doi.org/10.3390/bs13040297.

Fonseka, LN, & Woo, BKP 2023 "Therapeutic role of psilocybin and 3,4-methylenedioxymethamphetamine in trauma: A literature review", *World Journal of Psychiatry*, vol. 13, no. 5, pp. 182–190. https://doi.org/10.5498/wjp.v13.i5.182.

Friston, K 2013 "Life as we know it", *Journal of the Royal Society Interface*, vol. 10, no. 86, article 20130475. https://doi.org/10.1098%2Frsif.2013.0475

Friston, K 2010 "The free-energy principle: A unified brain theory?", *Nature Reviews. Neuroscience*, vol. 11, no. 2, pp. 127–38. https://doi.org/10.1038/nrn2787.

Friston, K, Friedman, DA, Constant, A, Knight, VB, Fields, C, Parr, T, & Campbell, JO 2023. "A variational synthesis of evolutionary and developmental dynamics", *Entropy* (Basel, Switzerland), vol. 25, no. 7, artcile 964. https://doi.org/10.3390/e25070964.

Friston, KJ, Lin, M, Frith, CD, Pezzulo, G, Allan Hobson, J, & Ondobaka, S 2017 "Active inference, curiosity and insight", *Neural Computation*, vol. 29, no. 10, pp. 2633–2683. https://doi.org/10.1162/neco_a_00999.

Friston, K, & Kiebel, S 2009 "Predictive coding under the free-energy principle", *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 364, no. 1521, pp. 1211–1221. https://doi.org/10.1098/rstb.2008.0300.

Friston, K, Kilner, J, & Harrison, L 2006 "A free energy principle for the brain", *Journal of Physiology* (Paris), vol. 100, no. 1–3, pp. 70–87. https://doi.org/10.1016/j.jphysparis.2006.10.001.

Friston, K, Thornton, C, & Clark, A 2012 "Free-energy minimization and the dark-room problem", *Frontiers in Psychology*, vol. 3, no. 130. https://doi.org/10.3389/fpsyg.2012.00130

Galilei, G 2005 [1623] *Il Saggiatore*. Besomi, O, & Helbing, M (eds.) Rome-Padua: Antenore.

Glick, D 2023 "The principle of least action and teleological explanation in physics", *Synthese* , vol. 202, no. 1, article 25. https://doi.org/10.1007/s11229-023-04251-x.

Guénin-Carlut, A 2023 "Strange things, statespace representation, and participatory realism comment on 'Path integrals, particular kinds, and strange things.' by Friston *et al.*", *Physics of Life Reviews*, vol. 47 (December), pp. 268–270. https://doi.org/10.1016/j.plrev.2023.10.027.

Guilielmo, B, & Mudry, L 2021 "Suspension du jugement (GP)". In: *Encyclopédie Philosophique*, https://Encyclo-Philo.Fr/.

Hanson, A 2021 ''Spontaneous electrical low-frequency oscillations: A possible role in Hydra and all living systems", *Philosophical Transactions of the Royal Society. Series B: Biological Sciences*, vol. 376, no. 1820, article 20190763. https://doi.org/10.1098/rstb.2019.0763

Ho, SS, Nakamura, Y, & Swain, JE 2020. "Compassion as an intervention to attune to universal suffering of self and others in conflicts: A translational framework", *Frontiers in Psychology*, vol. 11, artcile 603385. https://doi.org/10.3389/fpsyg.2020.603385.

von Uexküll, J 1965. *Animal Worlds and the Human World*. Paris: Denoël.

Kirchhoff, M, Parr, T, Palacios, E, Friston, K, & Kiverstein, J 2018 "The Markov blankets of life: Autonomy, active inference and the free energy principle", *Journal of the Royal Society, Interface*, vol. 15, no. 138, article 20170792. https://doi.org/10.1098/rsif.2017.0792.

Kiverstein, J, Kirchhoff, MD, & Froese, T 2022. "The problem of meaning: The free energy principle and artificial agency", *Frontiers in Neurorobotics*, vol. 16 (June), article 844773. https://doi.org/10.3389/fnbot.2022.844773.

Knill, DC, & Pouget. A 2004 "The Bayesian brain: The role of uncertainty in neural coding and computation", *Trends in Neurosciences*, vol. 27, no. 12, pp. 712–719. https://doi.org/10.1016/j.tins.2004.10.007.

Kuchling, F, Friston, K, Georgiev, G, & Levin, M 2020 "Morphogenesis as Bayesian inference: A variational approach to pattern formation and control in complex biological systems", *Physics of Life Reviews*, vol. 33 (July), pp. 88–108. https://doi.org/10.1016/j.plrev.2019.06.001.

Lesne, A 2014 "Shannon entropy: A rigorous notion at the crossroads between probability, information theory, dynamical systems and statistical physics", *Mathematical Structures in Computer Science*, vol. 24, no. 3, article e240311. https://doi.org/10.1017/S0960129512000783.

Longo, G 2020. "Information at the threshold of interpretation: Science as human construction of sense", *A Critical Reflection on Automated Science*, vol. 67. https://link.springer.com/chapter/10.1007/978-3-030-25001-0_5

Longo, G, Miquel, PA, Sonnenschein, C, & Soto, AM 2012 "Is information a proper observable for biological organization?", *Progress in Biophysics and Molecular Biology*, vol. 109, no. 3, pp. 108–114. https://doi.org/10.1016/j.pbiomolbio.2012.06.004.

Longo, G, & Montévil, M 2011 "From physics to biology by extending criticality and symmetry breakings", *Progress in Biophysics and Molecular Biology, Systems Biology and Cancer*, vol. 106, no. 2, pp. 340–347. https://doi.org/10.1016/j.pbiomolbio.2011.03.005.

Longo, G, & Montévil, M 2012 "Randomness increases order in biological evolution." In: Dinneen, MJ, Khoussainov, B, & Nies, A (eds.) *Computation, Physics and Beyond. WTCS 2012. Lecture Notes in Computer Science*, vol. 7160. Berlin-Heidelberg: Springer. https://doi.org/10.1007/978-3-642-27654-5_22

Longo, G, Montévil, M, & Kauffman, S 2012 "No entailing laws, but enablement in the evolution of the biosphere". In: *Genetic and Evolutionary Computation Conference*, pp. 1379–1392. New York: Acm.

Margulies, DS, Ghosh, SS, Goulas, A, Falkiewicz, M, Huntenburg, JM, Langs, G, & Bezgin. G 2016 "Situating the default-mode network along a principal gradient of macroscale cortical organization", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 44, pp. 12574–12579. https://doi.org/10.1073/pnas.1608282113.

Martin, WF, Garg, S, & Zimorski, V 2015 "Endosymbiotic theories for eukaryote origin", *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 370 no. 1678, article 20140330. https://doi.org/10.1098/rstb.2014.0330.

Mason, NL, Kuypers, KPC, Reckweg, JT, Müller, F., Tse, DHY, Da Rios, B, Toennes, SW, Stiers,P, Feilding, A, & Ramaekers, JG 2021 "Spontaneous and deliberate creative cognition during and after psilocybin exposure", *Translational Psychiatry*, vol. 11, no. 1, article 209. https://doi.org/10.1038/s41398-021-01335-5.

McGovern, HT, De Foe, A, Biddell, H, Leptourgos, P, Corlett, P, Bandara, K, & Hutchinson, BT 2022 "Learned uncertainty: The free energy principle in anxiety", *Frontiers in Psychology*, vol. 13, article 943785. https://doi.org/10.3389/fpsyg.2022.943785.

Menon, V 2023. "20 years of the default mode network: A review and synthesis", *Neuron*, vol. 111, no. 16, pp 2469–2487. https://doi.org/10.1016/j.neuron.2023.04.023.

Miquel, PA, & Hwang, SY 2016 "From physical to biological individuation", *Progress in Biophysics and Molecular Biology, From the Century of the Genome to the Century of the Organism: New Theoretical Approaches*, vol. 122, no. 1, pp. 51–57. https://doi.org/10.1016/j.pbiomolbio.2016.07.002.

Montévil, M 2019 "Measurement in biology is methodized by theory", *Biology & Philosophy*, vol. 34, no. 3, article 35. https://doi.org/10.1007/s10539-019-9687-x.

Montévil, M 2021 "Entropies and the Anthropocene crisis" *AI & Society*, May. https://doi.org/10.1007/s00146-021-01221-0.

Montévil, M 2022 "Disruption of biological processes in the Anthropocene: The case of phenological mismatch", *Acta Biotheoretica*, vol. 73, no. 2, article 5 https://hal-ens.archives-ouvertes.fr/hal-03574022.

Montévil, M, & Mossio, M 2015 "Biological organisation as closure of constraints", *Journal of Theoretical Biology*, vol. 372 (May), pp. 179–191. https://doi.org/10.1016/j.jtbi.2015.02.029.

Montévil, M, & Mossio, M 2020 "The identity of organisms in scientific practice: Integrating historical and relational conceptions", *Frontiers in Physiology*, vol. 11. https://www.frontiersin.org/article/10.3389/fphys.2020.00611.

Montévil, M 2019 "Possibility spaces and the notion of novelty: From music to biology", *Synthese*, vol. 196, no. 11, pp. 4555–4581. https://doi.org/10.1007/s11229-017-1668-5.

Montévil, M, Mossio, M, Pocheville, A, & Longo, G 2016a "Theoretical principles for biology: Variation", *Progress in Biophysics and Molecular Biology, From the Century of the Genome to the Century of the Organism: New Theoretical Approaches*, vol. 122, no. 1, pp. 36–50. https://doi.org/10.1016/j.pbiomolbio.2016.08.005.

Montévil, M, Speroni, L, Sonnenschein, C, & Soto, AM. 2016b. ''Modeling mammary organogenesis from biological first principles: Cells and their physical constraints", *Progress in Biophysics and Molecular Biology*, vol. 122, no. 1, pp. 58–69. https://doi.org/10.1016/j.pbiomolbio.2016.08.004

Moreno, A, & Mossio, M 2015. *Biological Autonomy: A Philosophical and Theoretical Enquiry*. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-9837-2.

Mossio, M, & Bich, L 2017 "What makes biological organisation teleological?", *Synthese*, vol. 194, no. 4, pp. 1089–1114. https://doi.org/10.1007/s11229-014-0594-z.

Mossio, M, Montévil, M, & Longo, G 2016 "Theoretical principles for biology: Organization", *Progress in Biophysics and Molecular Biology, From the Century of the Genome to the Century of the Organism: New Theoretical Approaches*, vol. 122, no. 1, pp. 24–35. https://doi.org/10.1016/j.pbiomolbio.2016.07.005.

Mossio, M, Saborido, C, & Moreno, A 2009 "An organizational account of biological functions", *The British Journal for the Philosophy of Science*, vol. 60, no. 4, pp. 813–841. https://doi.org/10.1093/bjps/axp036.

Nave, K 2025 *A Drive to Survive: The Free Energy Principle and the Meaning of Life*. Boston: MIT Press.

Palacios, ER, Razi, A, Parr, T, Kirchhoff, M, & Friston, K 2020 "On Markov blankets and hierarchical self-organisation", *Journal of Theoretical Biology*, vol. 486 (February), article 110089. https://doi.org/10.1016/j.jtbi.2019.110089.

Paperin, G, Green, DG, & Sadedin, S 2011 ''Dual-phase evolution in complex adaptive systems", *Journal of the Royal Society Interface*, vol. 8, no. 58, pp. 609-629. https://doi.org/10.1098/rsif.2010.0719

Parr, T, & Friston, KJ 2019 "Generalised free energy and active inference", *Biological Cybernetics*, vol. 113, no. 5–6, pp. 495–513. https://doi.org/10.1007/s00422-019-00805-w.

Ramstead, MJD 2023 "The free energy principle: A precise", *Dialectical Systems. A Forum for Biology, Ecology and Cognitive Science*. https://www.dialecticalsystems.eu/contributions/the-free-energy-principle-a-precis/

Ramstead, MJD, Kirchhoff, MD, Constant, A, & Friston, KJ 2021 "Multiscale integration: Beyond internalism and externalism", *Synthese*, vol. 198, no. 1, pp. 41–70. https://doi.org/10.1007/s11229-019-02115-x.

Ramstead, MJD, Sakthivadivel, DAR, Heins, C, Koudahl, M, Millidge, B, Da Costa, L, Klein, B, & Friston, KJ 2023 "On Bayesian mechanics: A physics of and by beliefs", *Interface Focus*, vol. 13, no. 3, article 20220029. https://doi.org/10.1098/rsfs.2022.0029.

Rayner, JG, Sturiale, SL, & Bailey, NW 2022 "The persistence and evolutionary consequences of vestigial behaviours", *Biological Reviews*, vol. 97, no. 4, pp. 1389–1407. https://doi.org/10.1111/brv.12847.

Rivera-García, MT, & Cruz, SL 2023 "The resurgence of hallucinogen drugs in clinical research", *Revista de Investigación Clínica*, vol. 75, no. 3, pp. 169–178. https://doi.org/10.24875/RIC.23000108.

Ruffini, G 2017 "An algorithmic information theory of consciousness", *Neuroscience of Consciousness*, no. 1, article nix019. https://doi.org/10.1093/nc/nix019.

Safron, A, Juliani, A, Reggente, N, Klimaj, V, & Johnson, M 2020 ''On the varieties of conscious experiences: Altered beliefs under psychedelics (ALBUS)". *Neuroscience of Consciousness*. PsyArXiv preprint. https://doi.org/10.31234/osf.io/zqh4b

Safron, A, Klimaj, V, & Hipólito, I 2022. ''On the importance of being flexible: Dynamic brain networks and their potential functional significances", *Frontiers in Systems Neuroscience*, vol. 15, article 688424. https://doi.org/10.3389/fnsys.2021.688424

Sánchez-Cañizares, J 2021. "The free energy principle: Good science and questionable philosophy in a grand unifying theory", *Entropy* (Basel, Switzerland), vol. 23, no. 2, article 238. https://doi.org/10.3390/e23020238.

Sarti, A, Citti, G, & Piotrowski, D 2019 "Differential heterogenesis and the emergence of semiotic function", *Semiotica*, vol. 2019, no. 230, pp. 1–34. https://doi.org/10.1515/sem-2018-0109.

Sarti, A, Citti, G, & Piotrowski, D 2022 "Expression and semiogenesis". In: Sarti, A, Citti, G, & Piotrowski, D (eds.) *Differential Heterogenesis: Mutant Forms, Sensitive Bodies*, pp. 131–192. Lecture Notes in Morphogenesis. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-97797-9_6.

Schrödinger, E 1944 *What Is Life? The Physical Aspect of the Living Cell*. Cambridge: Cambridge University Press.

Sharma, A, Czégel, D, Lachmann, M, Kempes, CP, Walker, SI, & Cronin, L 2023 ''Assembly theory explains and quantifies selection and evolution", *Nature*, vol. 622, no. 7982, pp. 321–328. https://doi.org/10.1038/s41586-023-06600-9

Slijepcevic, P 2024 "Principles of cognitive biology and the concept of biocivilisations", *BioSystems*, vol. 235 (January), artcile 105109. https://doi.org/10.1016/j.biosystems.2023.105109.

Smallwood, J, Bernhardt, BC, Leech, R, Bzdok, D, Jefferies, E, & Margulies, DS 2021 "The default mode network in cognition: A topographical perspective", *Nature Reviews. Neuroscience*, vol. 22, no. 8, pp. 503–513. https://doi.org/10.1038/s41583-021-00474-4.

Sonnenschein, C, & Soto, A 1999 *The Society of Cells: Cancer and Control of Cell Proliferation*. Oxford - New York: Bios Scientific Publishers - Springer. https://archive.org/details/societyofcellsca0000sonn

Soto, AM, Longo, G, Miquel, PA, Montevil, M, Mossio, M, Perret, N, Pocheville, A, & Sonnenschein, C 2016 "Toward a theory of organisms: Three founding principles in search of a useful integration", *Progress in Biophysics and Molecular Biology, From the Century of the Genome to the Century of the Organism: New Theoretical Approaches*, vol. 122, no. 1, pp. 77–82. https://doi.org/10.1016/j.pbiomolbio.2016.07.006.

Soto, AM, Longo, G, Montévil, M, & Sonnenschein, C 2016 "The biological default state of cell proliferation with variation and motility, a fundamental principle for a theory of organisms", *Progress in Biophysics and Molecular Biology, From the Century of the Genome to the Century of the Organism: New Theoretical Approaches*, vol. 122, no. 1, pp. 16–23. https://doi.org/10.1016/j.pbiomolbio.2016.06.006.

Stiegler, B 2021 *Nietzsche et la vie*. Paris: Gallimard.

Tahar, M 2023 "Agency, inventiveness, and animal play: Novel insights into the active role of organisms in evolution", *Spontaneous Generations*, vol. 11, no. 1. https://philarchive.org/rec/TAHAIA.

Tenaillon, O, Taddei, F, Radmian, M, & Matic, I 2001 "Second-order selection in bacterial evolution: Selection acting on mutation and recombination rates in the course of adaptation", *Research in Microbiology*, vol. 152, no. 1, pp. 11–16. https://doi.org/10.1016/s0923-2508(00)01163-3.

Timmermann, C, Kettner, H, Letheby, C, Roseman,L, Rosas, FE, & Carhart-Harris, RL 2021 "Psychedelics alter metaphysical beliefs", *Scientific Reports*, vol. 11, no. 1, article 22166. https://doi.org/10.1038/s41598-021-01209-2.

Umerez, J, & Mossio, M 2013 "Constraint", In: Dubitzky, W, Wolkenhauer, O, Cho, KH, Yokota, H (eds.) *Encyclopedia of Systems Biology*. New York: Springer. https://doi.org/10.1007/978-1-4419-9863-7_56.

Walker, SI, & Davies, PCW 2013 "The algorithmic origins of life", *Journal of the Royal Society, Interface*, vol. 10, no. 79, article 20120869. https://doi.org/10.1098/rsif.2012.0869.

Wang, L, & Smith, K 1998 "On chaotic simulated annealing", *IEEE Transactions on Neural Networks*, vol. 9, no. 4, pp. 716–718. https://doi.org/10.1109/72.701185.

Whinkin, E, Opalka, M, Watters, C, Jaffe, A, & Aggarwal, S 2023 "Psilocybin in palliative care: An update", *Current Geriatrics Reports*, vol. 12, no. 2, pp. 50–59. https://doi.org/10.1007/s13670-023-00383-7.

Zafar, R, Siegel, M, Harding, R, Barba, T, Agnorelli, C, Suseelan, S, Roseman, L, Wall, M, Nutt, DJ, & Erritzoe, D 2023 "Psychedelic therapy in the treatment of addiction: The past, present and future", *Frontiers in Psychiatry*, vol. 14, article 1183740. https://doi.org/10.3389/fpsyt.2023.1183740.

Organisms

SAPIENZA
UNIVERSITÀ DI ROMA

Special Issue, What AI Can Learn from Biology

Vol. 8, No. 1–2 (2025)
ISSN: 2532-5876

DOI: 10.13133/2532-5876/18627

The Rate of Entropy Production as a Lyapunov Function in Biophysical-chemical Systems

José Manuel Nieto-Villar,[a] *Mariano Bizzarri*[b] *and Ricardo Mansilla* [c, d*]

[a] *Department of Physical-Chemistry, A. Alzola Group of Thermodynamics of Complex Systems of M.V. Lomonosov Chair, Faculty of Chemistry, University of Havana, Cuba*
[b] *Department of Experimental Medicine, Sapienza University of Rome, via A. Scarpa 16-00161 Rome, Italy*
[c] *Centro Peninsular en Humanidades y Ciencias Sociales, CEPHCIS, UNAM, México*
[d] *Centro de Investigaciones Interdisciplinarias en Ciencias y Humanidades, UNAM, México*

***Corresponding author:** José Manuel Nieto-Villar, Email: nieto@fq.uh.cu

Abstract

An overview of the link between nonequilibrium thermodynamics and complexity theory is offered here, showing how the entropy production rate can be quantified through the spectrum of the Lyapunov exponents. The work shows how the entropy production per unit of time meets the necessary and sufficient conditions to be a Lyapunov function and constitutes *per se* an extremal principle. The entropy production fractal dimension conjecture is also established. The work demonstrates how the rate of entropy production as a non-extremal criterion represents an alternative way for sensitivity analysis of differential equations. Finally, in an extension to biophysical-chemical systems, on the one hand, the study presents the use of the dissipation function as a thermodynamic potential out of equilibrium in the characterization of biological phase transitions. On the other hand, it evidences that the entropy production rate represents a physical quantity that can be used to evaluate the complexity and robustness of cancer.

Keywords: nonequilibrium thermodynamics, entropy production rate, Lyapunov function, complexity, biological phase transitions

Citation: Nieto-Villar, JM, Bizzarri, M, & Mansilla, R 2025, "The Rate of Entropy Production as a Lyapunov Function in Biophysical-chemical Systems", *Organisms. Journal of Biological Sciences*, vol. 8, no. 1–2, pp. 85–96. DOI: 10.13133/2532-5876/18627

Introduction

The advent of the so-called chaos theory initially (Schuster, 2006) and the more recent developments in the sciences of complexity (Nicolis & Nicolis 2007) have drastically changed the vision of science, particularly the thermodynamics of irreversible processes.

The linear region of irreversible processes lies on a well-consolidated theory (Prigogine 1947; De Groot & Mazur 1962; Katchalsky & Curran 1965). However, the non-linear region is still waiting for a formalism to be built, on the one hand, while on the other, such a formalism also should incorporate complex phenomena. A first approximation in this direction, linking the thermodynamics of irreversible processes with nonlinear dynamics, was elaborated in the seminal work of Prigogine and colleagues (Nicolis & Prigogine 1977) under the name "dissipative structures." Beck & Schlögl published the work "Thermodynamics of chaotic systems" in the 1990s (Beck & Schlögl 1993), approach the subject. Although still far from a finished formalism, these works undertook the first steps in such a direction.

An extensive list of works in the literature addresses the relationship between nonequilibrium thermodynamics and complex phenomena (Gaspard *et al.* 2007; Nicolis & De Decker 2017; Nicolis & Nicolis 2010). A thermodynamic formalism of complex phenomena should be able to answer three fundamental aspects: 1. Formulate extremal principle for complex phenomena on a macroscopic scale; 2. Establish methods to determine stability in nonequilibrium states; 3. Formalize criteria to characterize the complexity at the macroscopic level of natural systems.

This work aims to offer a unifying overview of the relationship between nonequilibrium thermodynamics and non-linear dynamics, which, even far from establishing a finished formalism, serves as a starting point for what could constitute the theoretical bases of the "thermodynamics of complex phenomena." The work is structured as follows: Section 1 summarizes the fundamental aspects of the formalism of the thermodynamics of irreversible processes in the linear region; Section 2 offers an overview of the advances between nonequilibrium thermodynamics and complex phenomena; Section 3 provides an extension to biophysical-chemical systems.

1. The Formalism of the Thermodynamics of Irreversible Processes in the Linear Region

The seminal works of Onsager (Onsager 1931), De Groot-Mazur (De Groot & Mazur 1962), and Prigogine (Prigogine 1947) established the bases of the thermodynamics of irreversible processes. This formalism was based on four fundamental pillars:

1. Accept as a fundamental postulate that the production of entropy per unit of time $\frac{\delta S_i}{dt}$, is positive definite, that is:

$$\frac{\delta S_i}{dt} \equiv \dot{S}_i \geq 0 \text{, (1)}$$

2. Validity of the Onsager reciprocity relations.

3. Fulfillment of the "local equilibrium" hypothesis.

4. The existence of linear relationships between flows and forces.

In this way, the fundamental expression of the Second Law can be generalized as

$$\frac{dS_S}{dt} = \frac{\delta S_e}{dt} + \frac{\delta S_i}{dt} \text{, (2)}$$

where $\frac{dS_S}{dt} \equiv \dot{S}_S$ is the entropy rate of the system, $\frac{\delta S_e}{dt} \equiv \dot{S}_e$ is the rate of entropy exchange with the surroundings or entropy flow, and $\frac{\delta S_i}{dt} \equiv \dot{S}_i$ is the rate of entropy production. The Eq. (2) can be rewritten as

$$\dot{S}_S = \dot{S}_e + \dot{S}_i \text{, (3)}$$

Thus, the evolution criterion can be generalized as: $\dot{S}_i > 0$, which constitutes one of the postulates on which the formalism of irreversible processes rests and the essence of the Second Law. Additionally, it gives a physical meaning to time, which has been coined in the literature as The Arrow of Time (Coveney & Highfields 1991).

Sometimes, it is convenient, as we will see later, to use, instead of the rate of entropy production, the so-called dissipation function introduced by Lord Rayleigh, $\Psi \equiv T\dot{S}_i$, since it converts the entropy production rate into an out-of-equilibrium thermodynamic potential.

Formally, the rate of production of entropy, $\dot{S}_i$, can be evaluated as

$$\dot{S}_i = \sum_k J_k X_k \text{ , (4)}$$

where, J_k represents generalized flows, e.g., heat flow, substance flow, etc., and X_k are the generalized forces, that is, the causes that give rise to the appearance of flows, temperature gradients, substances, etc.

A linear relationship can be established between the flows and the generalized forces, known as the phenomenological (De Groot & Mazur 1962), which was established empirically long before the formal structure of the thermodynamics of irreversible processes was established. Hence, we have

$$J_k = L_{kk} X_k \text{ , (5)}$$

where, L_{kk} is known as a direct phenomenological coefficient, for example, the coefficient of thermal conductivity, λ, diffusion coefficient, D, etc. The formal structure of the thermodynamics of irreversible linear processes is based on the existence of equality, Eq. (5), that is the validity of linear relationships between generalized forces and flows. When there is no such phenomenological relationship, we speak of the non-linear region. It is essential to highlight that linearity in dynamic systems should be distinct from the existence of the linear dependence between flows and generalized forces, Eq. (5).

Of great importance are the coupling or interference processes (Prigogine 1961), which are subject to the Curie Principle of symmetry (Prigogine 1961); for example, given any two processes that are coupled under the Curie Principle, such that

$$\begin{aligned} J_1 &= L_{11}X_1 + L_{12}X_2, \\ J_2 &= L_{21}X_1 + L_{22}X_2 \end{aligned} \text{ , (6)}$$

where L_{11}, L_{22} are the straight phenomenological coefficients and L_{12}, L_{21} are known as cross-phenomenological coefficients. As we mentioned previously, point 2, concerning the so-called Onsager Reciprocity Principle, it is true that,

$$L_{12} = L_{21} \text{ , (7)}$$

In other words, the so-called Onsager Reciprocity Principle (De Groot & Mazur 1962; Onsager 1931) establishes that whenever an appropriate choice is made for the flows J_k and the forces X_k, the matrix of phenomenological coefficients is symmetric. Thus, considering Eqs. (6, 7) and substituting them in Eq. (4), we have that the rate of production of entropy for the coupling is given by

$$\begin{aligned} \dot{S}_i &= L_{11}X_1^2 + \left(L_{12} + L_{21}\right)X_1X_2 + L_{22}X_2^2 \\ &= L_{11}X_1^2 + 2L_{12}X_1X_2 + L_{22}X_2^2 \ge 0. \end{aligned} \text{ , (8)}$$

The Eq. (8) is a semi-positive definite quadratic form by the Second Law. Linear algebra imposes restrictions on the phenomenological coefficients in formula (8); it must be true that

$$\begin{aligned} & L_{11} > 0, L_{22} > 0 \\ & \left(L_{12} + L_{21}\right)^2 < 4L_{11}L_{22}. \end{aligned} \text{ , (9)}$$

The straight coefficients are always positive magnitudes, while the crossed ones can take any value as long as the inequality of the last expression of Eq. (9).

The stationary states, also known as fixed points in the theory of dynamical systems (Andronov *et al.* 1966), are states through which different processes, physical, chemical, biological, etc. (De Groot & Mazur 1962; Katchalsky & Curran 1965) and are of particular interest in the framework of the theory of complexity sciences (Nicolis & Nicolis 2007).

Formally, a dynamical system can be defined as the ordered pair $\left(\mathrm{E}, \mathrm{T}_t\right)$ where E represents an appropriate manifold and T_t is a one-parameter group of diffeomorphisms under the parameter t often represented by time. If one has an atlas of local charts for the manifold E, on those charts, it is possible a representation the dynamical system in the following form: $\dot{X}(t) = F\left(X(t)\right)$, where F is the vector field associated with the one-parameter group of diffeomorphisms.

It is said that the solution $X(t) = X_0$ is an equilibrium position or a stationary state of the system if $F\left(X_0\right) = 0$. We further say that X_0 is an attractor of the system, if for any other solution $X(t)$, whose initial conditions are close enough to X_0, we have $X(t) \rightarrow X_0$ when $t \rightarrow \infty$.

From a nonequilibrium thermodynamics point of view (De Groot & Mazur 1962), a stationary state is formally defined as a dynamic state, for which it is true that during a finite time, the state variables and the control parameters remain constant, and dissipative flows are verified, that is to say $\dot{S}_i > 0$, in such a way that

$$\dot{S}_i = -\dot{S}_e, \quad (10)$$

That is, at the same rate that entropy is produced $\dot{S}_i$, exchanges with surroundings $\dot{S}_e$, in such a way that $\dot{S}_S = 0$. Furthermore, steady states are characterized by the number of forces k that remain constant; hence, the stationary states of an order made references to k (De Groot & Mazur 1962). For instance, in Eq. (8), assuming there is a steady state, for X_2 constant, that is, of order one, $k = 1$, we should have to verify Prigogine's Theorem of Minimum Entropy Production or Prigogine's Principle (Prigogine 1961), which ensures the stability of the stationary state, that is, out of equilibrium, which constitutes an extension of the stability criterion in the vicinity of the equilibrium, Gibbs-Duhem Principle (Kondepudi & Prigogine 1998). In this way, Prigogine's Principle represents, in fact, an extremal principle if the linear relationships between flows and forces are fulfilled—Eq. (5).

Glansdorff and Prigogine tried to generalize Prigogine's Principle, known as the "general criterion of evolution" (Glansdorff & Prigogine 1971), demonstrating how the rate of entropy production, Eq. (8), constitutes from physics, a natural Lyapunov function (Mawhin 1996). According to the procedure proposed by Glansdorff and Prigogine, the entropy production per unit of time $\dot{S}_i$ is identified as a Lyapunov function, $V(x)$, $\dot{S}_i \equiv V(x)$, such that

$$\begin{gathered} \dot{S}_i \equiv V(x) \geq 0, \\ \frac{d\dot{S}_i}{dt} \leq 0. \end{gathered} \quad , (11)$$

The Eulerian derivative of the entropy production rate, Eq. (4), is given by

$$\begin{aligned} \frac{d\dot{S}_i}{dt} &= \sum_k J_k \frac{dX_k}{dt} + \sum_k X_k \frac{dJ_k}{dt} \\ &= \frac{d_X(\dot{S}_i)}{dt} + \frac{d_J(\dot{S}_i)}{dt}; \end{aligned} \quad , (12)$$

Considering Eq. (8) and substituting in Eq. (12), one has:

$$\begin{aligned} \frac{d_X(\dot{S}_i)}{dt} &= J_1 \frac{dX_1}{dt} + J_2 \frac{dX_2}{dt}, \\ \frac{d_J(\dot{S}_i)}{dt} &= X_1 \frac{dJ_1}{dt} + X_2 \frac{dJ_2}{dt}; \end{aligned} \quad , (13)$$

Taking into account Eqs. (6), (7), and (13), and substituting in Eq. (12) is obtained

$$\begin{aligned} \frac{d(\dot{S}_i)}{dt} &= 2J_1 \frac{dX_1}{dt} + 2J_2 \frac{dX_2}{dt}, \\ &= 2\frac{d_X(\dot{S}_i)}{dt}, \end{aligned} \qquad (14)$$

$$\frac{1}{2}\frac{d(\dot{S}_i)}{dt} < 0;$$

In this way, it is demonstrated that formula (14), as the production of entropy per unit of time, is a physical magnitude that constitutes *per se* a Lyapunov function if there is a linear dependence between the flows and the generalized forces. As can be seen, the general criterion of evolution, formula (14), is restricted to the linear region of irreversible processes.

2. Thermodynamic Formalism of Complex Processes

As we commented at the beginning, unlike the formalism of the thermodynamics of irreversible processes in the linear region, where most of its precepts are consolidated, the nonlinear region is still in the making; due to this, it is still premature to speak of a finished formalism. That is why we intend to provide a landscape approach to the subject and, above all, try to articulate the thermodynamic formalism of irreversible processes with that of nonlinear dynamics so that it allows us to offer a thermodynamic approach to complex phenomena (Mansilla & Nieto-Villar 2017).

On the one hand, it is essential to be clear about what we refer to as complex (Bizzarri *et al.* 2020).

Hence, the complexity manifested by dynamical systems highlights the following general and critical aspects of understanding this phenomenon:

1. Complex should not be seen as synonymous with complicated since a system described by few degrees of freedom can exhibit high complexity during its evolution; on the contrary, a system that requires many degrees of freedom to be able to describe itself and which is therefore complicated, may or may not exhibit complex behavior.

2. Complexity manifests itself through the appearance of emergent properties. These are macroscopic observables that can only sometimes be deduced from the interaction rules that govern the evolution of the different components of the systems.

3. The dimension of the patterns, both temporal and spatial, is generally not an integer and is greater than its topological dimension; therefore, they are said to have a fractal dimension (Betancourt-Mar *et al.* 2016).

4. On many occasions, the complex processes described through deterministic dynamic systems show a sensitive dependence on the initial conditions. This behavior can be confused with stochastic processes and is known as deterministic chaos (Strogatz 2000). The most important consequence of this property is the impossibility of making predictions about the system's evolution in the long term. In other words, the so-called Laplacian determinism collapses.

5. For a deterministic dynamic system to exhibit complex behavior, it must meet two fundamental requirements: nonlinear and that feedback processes exist (Nieto-Villar *et al.* 2013).

6. The fundamental mechanism that describes a system's emergent properties and complexity is based on the occurrence of bifurcations (Nicolis 1972; Nicolis & Daems 1998), a dynamic analog of phase transitions. The bifurcations exhibit a universal character in their phenomenology (Kuznetsov 2013), making them independent of the system's characteristics and representing a source of innovation and diversification because they give systems a new type of solution. The fluctuations, which have a microscopic origin, grow and amplify until they reach the macroscopic level, which leads to a break in the spacetime symmetry, giving rise to self-organization outside of thermodynamic equilibrium, the establishment of order, and coherence on a macroscopic scale, and consequently to the appearance of complexity.

Hence, the term complex should not be seen as a synonym for complicated; that is, dynamic systems self-organize temporally and spatially out of thermodynamic equilibrium, a term coined by Prigogine as Dissipative Structures (Prigogine 1978), which gives rise to the manifestation of complex phenomena.

On the other hand, Seth Lloyd compiled an extensive, still incomplete list of ways to measure complexity (Lloyd 2001). This include Shannon, Gibbs-Boltzmann, Renyi, Tsallis, Kolmogorov-Sinai entropies, and fractal dimension.

Even today, there is a great controversy concerning the thermodynamic formalism of irreversible processes, including Prigogine's Principle of Entropy Production. According to Bruers (Bruers 2006), at least "six principles" can be mentioned: 1. Principle of minimum dissipation close to equilibrium; 2. Principle of minimum production of entropy near equilibrium; 3. Principle of maximum production of entropy near equilibrium; 4. Non-variational principle far from the equilibrium of maximum production of entropy; 5. Variational principle far from the equilibrium of maximum production of entropy; 6. Optimization of the principle of minimum production of entropy.

Chemical reactions constitute an ideal model to delve into the subject since, firstly, they can occur "close to or far" from thermodynamic equilibrium, and, secondly, there is no linear relationship between the generalized flow, the rate of reaction $\dot{\xi}$, and generalized force, an affinity for the inverse of temperature $\frac{1}{T}\mathcal{A}$. Furthermore, their dynamics exhibit a wide range of temporal and spatial complexity (Nieto-Villar & Velarde 2001), and the developed formalism can be extended to biological systems.

Briefly, we will show how it is possible to generalize, at least for chemical and biological processes, the "general criterion of evolution" of Glansdorff-Prigogine (Glansdorff & Prigogine 1971), demonstrating how the rate of entropy production is a Lyapunov function without the need for the linear relationships between flows and forces hold.

Lyapunov, in his 1892 doctoral thesis (Mawhin 1996), developed a mathematical method that allowed knowing the evolution and global stability of a dynamical system, known as the Lyapunov function $V(x)$ (see Fig. 1 in Appendix). Thus, we have succinctly that:

Let p be a fixed point, a steady state of a flow $\frac{dx}{dt} \equiv \dot{x} = f(x)$, such that, if for some neighborhood N of p the following conditions hold:

1. $V(x) > 0 \ \forall x \neq p$ in N and $V(p) = 0$;
2. The Eulerian derivative, $\frac{dV(x)}{dt} \leq 0$ for $\forall x$ in N.

The function $V(x)$ is called Lyapunov´s function. Thus, it can be stated that for all $t \geq t_0$, p is stable, and if $\frac{dV(x)}{dt} < 0$, the equilibrium position is asymptotically stable.

On the one hand, we show that the entropy production per unit time, at least for chemical reactions, meets the necessary and sufficient conditions of a Lyapunov function (Nieto-Villar *et al.* 2003) and, in fact, constitutes an extremal criterion *per se*, regardless of whether the network of chemical reactions is "near" or "far" from equilibrium. Recently, it has been demonstrated in reaction-diffusion-type systems (Ledesma-Durán & Santamaría-Holek 2022).

On the other hand, it was shown (Nieto-Villar *et al.* 1995; Garcia-Fernández *et al.* 1996; Nieto-Villar *et al.* 2013; Nieto-Villar *et al.* 2022) using an Ansatz through a functional of the rate of entropy production of the control parameters of the dynamic system, Ω, as

$$\dot{S}_i = f(\Omega) > 0 \text{, (15)}$$

Thus, it is found that the Eulerian derivative of Eq. (15) holds the following:

$$\frac{d\dot{S}_i}{dt} \equiv \Im = \frac{\partial \dot{S}_i}{\partial \Omega}\frac{d\Omega}{dt} \leq 0 \text{, (16)}$$

In this way, we have the acceleration of the production of entropy rate, $\Im \equiv \frac{d\dot{S}_i}{dt}$, which constitutes *per se* a potential function out of equilibrium.

The works of Hoover and Nose (Hoover & Posch 1994; Hoover 2007) and Gaspard (Gaspard 2007) showed that the rate of entropy production $\dot{S}_i$ is related to the spectrum of the Lyapunov exponents λ_j through the relationship,

$$\frac{dS_i}{dt} \equiv \dot{S}_i \approx -\sum_j \lambda_j > 0 \text{, (17)}$$

The formula, Eq. (17), establishes *per se* a natural link between the formalism of the thermodynamics of irreversible processes and nonlinear dynamics regardless of whether the system evolves "close" or "far" from thermodynamic equilibrium.

It is known that sensitivity analysis of differential equations has been used successfully to determine the fundamental steps in a reaction mechanism (Varma 2005). Edelson's pioneering works (Edelson & Allara 1980; Edelson & Thomas 1981; Edelson 1983) allowed the identification of the fundamental steps in a mechanism and its reduction. Later, Turanyi used the method in the famous Belousov-Zhabotinsky BZ reaction (Turányi 1990; Gyorgyi *et al.* 1990; Turányi 1993), drastically reducing the model mechanism, GTF, from 81 to 42 steps.

As an alternative method to the sensitivity analysis, we proposed using the entropy production rate as a non-extremal criterion, called the Method of Dominant Steps (Nieto-Villar & Velarde 2001; Nieto-Villar *et al.* 2022; Rieumont-Briones *et al.* 1997). For this, we postulate that those steps that exhibit a greater value of entropy production would be the fundamental ones in a reaction mechanism for fixed values of the control parameters.

Let be a mechanism of reaction composed of n-reaction steps and m-species, represented by equality (18), as

$$\begin{aligned} x_{1/i} &= x_{2/i} \\ &\vdots \qquad \text{, (18)} \\ x_{m-1/n} &= x_{m/n} \end{aligned}$$

Thus, we have that the rate of production of entropy of the step-n is given by

$$\dot{S}_{i/n} = R\left(\dot{\xi}_{+/n} - \dot{\xi}_{-/n}\right) \ln \frac{\dot{\xi}_{+/n}}{\dot{\xi}_{-/n}} \geq 0 \text{, (19)}$$

where $\dot{\xi}_{+/n}$, $\dot{\xi}_{-/n}$ are forward a reverse chemical rate of the step-n. Step n will be dominant compared to step n-1 if it is fulfilled that: $\dot{S}_{i/n} > \dot{S}_{i/n-1}$. In this way, the rate of entropy production, as a non-extremal criterion, generalizes the so-called "maximum entropy" criterion later proposed by Martyushev and Seleznev (Martyushev & Seleznev 2006) and constitutes a complementary method to the sensitivity analysis of differential equations.

The fractal dimension D_f represents one of the most important properties of an attractor of a dynamic system and a way to estimate the complexity of spatiotemporal patterns from the geometric point of view (Farmer 1982), as we mentioned at the beginning of this section. Grassberger (Grassberger & Procaccia 1983) proposed a generalization of the fractal dimension, the generalized fractal dimension D_q as

$$D_q = \lim_{\varepsilon \to \infty} \frac{S_q(R)}{\ln\left(\frac{1}{\varepsilon}\right)}, \text{(20)}$$

where $S_q(R)$ is the Renyi´s entropy (Rényi 1960). From the formula, Eq. (20), three basic dimensions are obtained as particular cases: D_0, D_1, D_2; the Hausdorff-Besicovitch fractal dimension D_0, the informational dimension (Farmer 1982), $D_1 = \lim_{q \to 1} D_q$, and the correlation dimension D_2. In the case of fractals, the three dimensions are approximately equal, while in multifractals, it is true that: $D_0 > D_1 > D_2$ (Farmer 1983).

An alternative and straightforward way to compute the fractal dimension of a dynamical system is through the spectrum of Lyapunov exponents. λ_j, known as the Lyapunov dimension D_L defined through the Kaplan-York conjecture (Frederickson 1983) as:

$$D_L = j + \frac{\sum_{i=1}^{j} \lambda_i}{\left|\lambda_{j+1}\right|}, \text{(21)}$$

where j is the largest integer for which it is true that: $\lambda_1 + \lambda_2 + \cdots + \lambda_j \geq 0$. By analogy to Eq. (21), we established through an ansatz the following conjecture: the fractal dimension of entropy production (Betancourt-Mar *et al.* 2016), defined as:

$$D_{\dot{S}_i} = j + \frac{\dot{S}_i}{\left(\sum_{i=j+1}^{n} \lambda_i\right)}, \text{(22)}$$

where the entropy production per unit time $\dot{S}_i$, is evaluated through the formula (17), n is the number of all Lyapunov exponents.

3. Extension to Biophysical-Chemical Systems

Finally, we will provide a brief landscape of the application of the thermodynamic formalism of complex processes in biological systems, particularly on the topic of the emergence and evolution of cancer. Non-equilibrium thermodynamics has been successfully used in studies of longevity, aging, the origin of life, and, in particular, cancer (Miquel *et al.* 1984; Balmer 1982; Nieto-Villar *et al.* 2003; Molnar *et al.* 2005; Luo 2009; Lucia 2014; Lucia *et al.* 2015; Marin & Sabater 2017; Triana *et al.* 2018; Betancourt-Mar *et al.* 2018; Montemayor-Aldrete *et al.* 2020; Mesa-Rodríguez *et al.* 2022; Michaelian 2022; Nieto-Villar & Mansilla 2022; Miranda & Souza 2023).

We must start with a formal definition: ...cancer is a complex network of cells that have lost their specialization and control of growth, and that appears through a "biological phase transition" leading to spatiotemporal self-organization outside the thermodynamic equilibrium. This exhibits high robustness, adaptability, complexity, and hierarchy, which enables the creation of new information and learning capacity (Montero *et al.* 2018).

The diagnosis of the proliferative and invasive capacity of a tumor is a complicated issue since these terms include many factors. Let us highlight two fundamental ones: aggressiveness, which is related to the speed of tumor growth, and malignancy, the ability of the tumor to invade and infiltrate healthy tissue, associated with its morphological characteristics (roughness) (Norton 2005).

The growth rate of the tumor, ξ is given by

$$\xi = \xi_m - \xi_{ap}, \text{(23)}$$

where ξ_m, ξ_{ap} are the rates of mitosis (cell division) and apoptosis (programmed cell death), respectively. By analogy to Eq. (19), we can evaluate the production of entropy per unit of time $\dot{S}_i$, during the growth of a tumor (Izquierdo-Kulich *et al.* 2011) as

$$\dot{S}_i = \left(\dot{\xi}_m - \dot{\xi}_{ap}\right)\ln\frac{\dot{\xi}_m}{\dot{\xi}_{ap}} \geq 0, \quad (24)$$

On the other hand, we developed a method based on knowing the rates of mitosis $\dot{\xi}_m$ and apoptosis $\dot{\xi}_{ap}$ (Izquierdo-Kulich & Nieto-Villar 2013) to quantify morphological characteristics (roughness) of the tumor, the malignancy of a tumor, through the fractal dimension D_f, as

$$D_f = \left(\frac{5\dot{\xi}_{ap} - \dot{\xi}_m}{\dot{\xi}_m + \dot{\xi}_{ap}}\right), \quad (25)$$

Considering Eqs. (23) and (25), we can rewrite Eq. (24) depending on the rate of tumor growth, $\dot{\xi}$ and the fractal dimension of the tumor D_f as

$$\dot{S}_i = R\dot{\xi}\ln\left(\frac{5 - D_f}{1 + D_f}\right), \quad (26)$$

In this way, an appropriate expression is obtained, Eq. (26), to evaluate the production of entropy per unit of time $\dot{S}_i$, during the emergence and evolution of cancer, which relates to two fundamental properties of tumors: aggressiveness and malignancy (Izquierdo-Kulich *et al.* 2011). Thus, we can affirm that the production of entropy per unit of time represents a physical quantity to evaluate cancer's complexity as well as robustness, namely the ability of a system to continue functioning in the face of internal or external perturbations or fluctuations.

Landau's seminal work (Landau & Lifshitz 1964) proposed a theory of continuous phase transitions in which symmetry breaking occurs near the critical point. In correspondence with the formalism proposed by Landau, a potential function is defined Φ, known as the Landau potential. The Landau potential Φ is defined in terms of the state variables that characterize the system, for example, temperature and pressure, as well as a function of the so-called order parameter η, which is empirically defined.

To formalize out-of-equilibrium phase transitions, a term we coined as biological phase transition (Betancourt-Mar *et al.* 2017), during the emergence and evolution of cancer, we selected the dissipation function, $\Psi \equiv T\dot{S}_i$, which is a non-equilibrium thermodynamic potential as an analogy to the Landau potential Φ.

Thus, we have that, in the case of the emergence and evolution of cancer, biological phase transition is selected as an order parameter η, the difference between the fractal dimension of healthy cells D_f^H and the fractal dimension of tumor cells D_f^T, such that:

$$\eta = D_f^H - D_f^T, \quad (27)$$

Thus, we have that at the critical point P_C it holds that $\eta = 0$ and so on in any other "ordered" phase $\eta \neq 0$. In this way, the order parameter η is called the degree of complexity (Betancourt-Mar *et al.* 2017).

Considering Eqs. (27) and (26), and making a power series expansion of the dissipation function Ψ, assuming for simplicity that $D_f^H = 1$, is obtained

$$\Psi\left(\dot{\xi}, d_f^C\right) = \Psi_0\left(\dot{\xi}, d_f^C\right) + \alpha\left(\dot{\xi}, d_f^C\right)\eta^2 + \beta\left(\dot{\xi}, d_f^C\right)\eta^4, \quad (28)$$

Eq. (28) represents an out-of-equilibrium extension of Landau's Theory and allows formalizing biological phase transitions through non-equilibrium thermodynamics. In this way, we understand how the development of a primary tumor from a microscopic level—an avascular growth—to a macroscopic level—the vascular phase—and the subsequent appearance of metastases do not occur simply by accumulation of malignant cells but through bifurcations, i.e., a biological phase transition (Izquierdo-Kulich *et al.* 2013; Llanos-Pérez *et al.* 2015; Llanos-Pérez *et al.* 2016; Martin *et al.* 2017; Betancourt-Mar *et al.* 2017; Guerra, A, *et al.* 2018; Betancourt-Padron *et al.* 2020; Nieto-Villar & Mansilla 2021).

Conclusions and Remarks

In summary, non-equilibrium thermodynamics and nonlinear dynamics articulate coherently. This let us establish a formal path of what could become the thermodynamics of complex processes. As essential aspects, it was shown that:

1. On the one hand, the entropy production rate is a physical magnitude representing a Lyapunov function *per se*, regardless of whether the dynamic system is close to or far from equilibrium, constituting an extremal criterion.
2. Conversely, the entropy production rate constitutes a complementary method to the sensitivity analysis of differential equations and appears as a non-extremal criterion.
3. An extension of the formalism to biophysical-chemical systems, on the one hand, shows the use of the dissipation function as a non-equilibrium thermodynamic potential in the characterization of biological phase transitions.
4. On the other hand, it was evidenced that the rate of entropy production represents a physical magnitude useful to evaluate the complexity and robustness of cancer and it may be used as a quantitative index of the metastatic potential of tumors.

Appendix: Lyapunov Function

Let the dynamical system be defined by:

$$\begin{bmatrix} \frac{dx_1}{dt} \\ \vdots \\ \frac{dx_n}{dt} \end{bmatrix} = \begin{bmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_n(x_1, \dots, x_n) \end{bmatrix} \qquad (1)$$

where the functions $f_1, \dots f_n$ are assumed to be continuous and have continuous first-order partial derivatives with respect to all variables $x_1, \dots, x_n$. Let us further suppose that:

$$f_i(0, \dots, 0) = 0 \quad ; \quad i = 1, \dots n$$

That is, the origin of the coordinates $(0, \dots, 0)$ is an equilibrium position of the system.

It is said that the function $V(x_1, \dots x_n)$ is a Lyapunov function for the equilibrium position of the system if:

a) $V(x_1, \dots x_n)$ is continuous in a neighborhood $\mathfrak{V}$ of the point $(0, \dots, 0)$, as well as all its first-order derivatives with respect to the variables $x_1, \dots, x_n$. Further:

$$V(x_1, \dots x_n) \geq 0$$

in the neighborhood $\mathfrak{V}$ of the point $(0, \dots, 0)$.

b) The derivative concerning the system (1):

$$\frac{d}{dt} V\big(x_1(t), \dots x_n(t)\big) = \sum_{i=1}^{n} \frac{\partial}{\partial x_i} V\big(x_1(t), \dots x_n(t)\big) \frac{d}{dt} x_1(t) < 0$$

where $\big(x_1(t), \dots x_n(t)\big)$ is a trajectory of the system (1).

Notice that:

$$\sum_{i=1}^{n} \frac{\partial}{\partial x_i} V\big(x_1(t), \dots x_n(t)\big) \frac{d}{dt} x_1(t) = \langle grad\, V\big(x_1(t), \dots x_n(t)\big), [f_1, \dots f_n] \rangle$$

That is, the scalar product of the gradient of the function $V(x_1, \dots x_n)$ evaluated in the trajectory $\big(x_1(t), \dots x_n(t)\big)$ of the system and the vector field of the system evaluated in the same trajectory:

$$[f_1, \dots f_n] = \big[f_1\big(x_1(t), \dots, x_n(t)\big), \dots, f_n\big(x_1(t), \dots, x_n(t)\big)\big]$$

The fact that this scalar product is less than zero indicates that the angle between the vector $gradV$ and $[f_1, \dots f_n]$ must be bigger than 90°. This condition guarantees the asymptotic stability of the equilibrium position $(0, \dots, 0)$.

Figure 1 shows what was previously described for the case of $n = 2$.

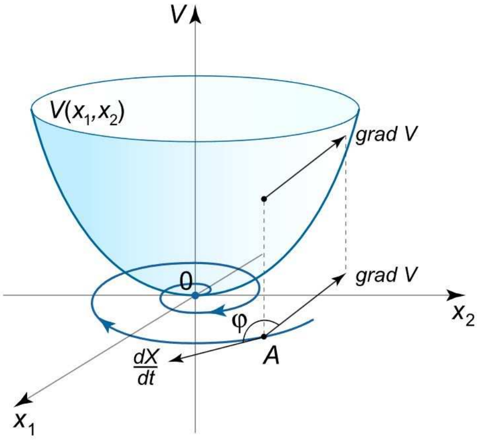


Figure 1

Acknowledgments

Prof. Dr. Germinal Cocho and Prof. Dr. A. Alzola *in memoriam*. One of the authors (JMNV) thanked the CEPHCIS of UNAM Mexico, for the warm hospitality and the financial support by PREI-DGAPA-UNAM-2022. Finally, we thank the anonymous reviewers for their helpful comments and interesting suggestions.

References

Andronov, A, Vit, A, & Chaitin, C 1966 *Theory of Oscillators*. Oxford: Pergamon Press.

Aoki, I 1991, "Entropy principle for human development, growth and aging", *Journal of Theoretical Biology*, vol. 150, no. 2, pp. 215–223.

Balmer, RT 1982 "Entropy and aging in biological systems", *Chemical Engineering Communications*, vol. 17, no.1-6, pp. 171-181.

Beck, C, & Schlögl, F 1993 *Thermodynamics of Chaotic Systems: An Introduction*. New York: Cambridge University Press.

Betancourt-Mar, JA, Rodríguez-Ricard, M, Mansilla, R, Cocho, G, & Nieto-Villar, JM 2016 "Entropy production: Evolution criteria, robustness and fractal dimension", *Revista Mexicana de Física*, vol. 62, no. 2, pp. 164–167.

Betancourt-Mar, JA, Llanos-Pérez, JA, Cocho, G, Mansilla, R, Martin, RR, Montero, S, & Nieto-Villar, JM 2017 "Phase transitions in tumor growth: IV relationship between metabolic rate and fractal dimension of human tumor cells", *Physica A: Statistical Mechanics and its Applications*, vol. 473, pp. 344–351.

Betancourt-Mar, JA, Mansilla, R, Cocho, G, & Nieto-Villar, JM 2018 "On the relationship between aging & cancer", *MOJ Gerontology & Geriatrics*, vol. 3, no. 2, pp. 163–168.

Betancourt-Padron, PJ, García-Medina, K, Mansilla, R, & Nieto-Villar, JM 2020 "Phase transition in tumor growth VIII: The spatiotemporal avascular evolution", *Revista Mexicana de Física*, vol. 66, no. 6, pp. 856–862.

Bizzarri, M, *et al.* 2020 "Complexity in biological organization: Deconstruction (and subsequent restating) of key concepts", *Entropy*, vol. 22, no. 8, p. 885.

Bruers, S 2006 "Classification and discussion of macroscopic entropy production principles", *arXiv* preprint cond-mat/0604482.

Coveney, P & Highfields, R 1991 *The Arrow of Time: A Voyage Through Science to Solve Time's Greatest Mystery*. Columbine: Fawcett (1st edition).

De Groot, SR & Mazur, P 1962 *Non-Equilibrium Thermodynamics*. Amsterdam: North-Holland Publishing Company.

Edelson, D, & Allara, DL 1980 "A computational analysis of the alkane pyrolysis mechanism: Sensitivity analysis of individual reaction steps", *International Journal of Chemical Kinetics*, vol. 12, no. 9, pp. 605–621.

Edelson, D, & Thomas, VM 1981 "Sensitivity analysis of oscillating reactions. 1. The period of the Oregonator", *The Journal of Physical Chemistry*, vol. 85, no. 11, pp. 1555–1558.

Edelson, D 1983 "Sensitivity analysis of proposed mechanisms for the Briggs-Rauscher oscillating reaction", *The Journal of Physical Chemistry*, vol. 87, no. 7, pp. 1204–1208.

Farmer, JD 1982 "Dimension, fractal measures, and chaotic dynamics. In: Haken, H. (ed) *Evolution of Order and Chaos*. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-68808-9_20.

Farmer, JD 1982, "Information dimension and the probabilistic structure of chaos", *Zeitschrift für Naturforschung A*, vol. 37, no. 11, pp. 1304–1326.

Farmer, JD, Ott, E, & Yorke, JA 1983 "The dimension of chaotic attractors", *Physica D: Nonlinear Phenomena*, vol. 7, no. 1–3, pp. 153–180.

Frederickson, P, Kaplan, JL, Yorke, ED, & Yorke, JA 1983 "The Liapunov dimension of strange attractors", *Journal of Differential Equations*, vol. 49, no. 2, pp. 185–207.

Garcia-Fernández, JM, Nieto-Villar, JM, & Rieumont-Briones, J 1996 "The rate of entropy production as an evolution criterion in chemical systems II. Chaotic reactions", *Physica Scripta*, vol. 53, no. 6, pp. 643–644.

Gaspard, P, Henneaux, M, & Lambert, F (eds.) 2007 *From Dynamical Systems Theory to Nonequilibrium Thermodynamics. Symposium Henri Poincaré, Proceedings*. Brussels: International Solvay Institutes for Physics and Chemistry.

Gaspard, P 2007 "Time asymmetry in nonequilibrium statistical mechanics", *Advances in Chemical Physics*, vol. 135, pp. 83–134.

Glansdorff, P, & Prigogine, I 1971 *Thermodynamics of Structure, Stability and Fluctuations*. New York: Wiley.

Grassberger, P, & Procaccia, I 1983, "Characterization of strange attractors", *Physical Review Letters*, vol. 50, no. 5, pp. 346–349.

Guerra, A, *et al.* 2018 "Phase transitions in tumor growth VI: Epithelial–Mesenchymal transition", *Physica A: Statistical Mechanics and its Applications*, vol. 499, pp. 208–215.

Gyorgyi, L, Turányi, T, & Field, RJ 1990 "Mechanistic details of the oscillatory Belousov-Zhabotinskii reaction", *Journal of Physical Chemistry*, vol. 94, no. 18, pp. 7162–7170.

Hoover, WG, & Posch, HA 1994 "Second-law irreversibility and phase-space dimensionality loss from time-reversible nonequilibrium steady-state Lyapunov spectra", *Physical Review E*, vol. 49, no. 3, pp. 1913–1920.

Hoover, WG 2007 "Nosé–Hoover nonequilibrium dynamics and statistical mechanics", *Molecular Simulation*, vol. 33, no. 1–2, pp. 13–19.

Izquierdo-Kulich, E, Alonso-Becerra, E, & Nieto-Villar, JM 2011 "Entropy production rate for avascular tumor growth", *Journal of Modern Physics*, vol. 2, no. 06, pp. 615–620.

Izquierdo-Kulich, E, & Nieto-Villar, JM 2013 "Morphogenesis and complexity of the tumor patterns". In: Rubio, R. *et al.* (eds.) *Without Bounds: A Scientific Canvas of Nonlinearity and Complex Dynamics. Understanding Complex Systems*. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-34070-3_48.

Izquierdo-Kulich, E, Rebelo, I, Tejera, E, & Nieto-Villar, JM 2013 "Phase transition in tumor growth: I avascular development", *Physica A: Statistical Mechanics and its Applications*, vol. 392, pp. 6616–6623.

Katchalsky, A, & Curran, P 1965 *Non-Equilibrium Thermodynamics in Biophysics*. Cambridge: Harvard University Press.

Kondepudi, D & Prigogine, I 1998 *Modern Thermodynamics, From Heat Engines to Dissipative Structures*. New York: John Wiley & Sons.

Kuznetsov, YA 2013 *Elements of Applied Bifurcation Theory*. Cham: Springer.

Landau, LD, & Lifshitz, EM 1964 *Curso de Física Teórica, Física Estadística*, Vol. 5. México: Reverté.

Ledesma-Durán, A, & Santamaría-Holek, I 2022 "Energy and Entropy in Open and Irreversible Chemical Reaction–Diffusion Systems with Asymptotic Stability", *Journal of Non-Equilibrium Thermodynamics*, vol. 47, no. 3, pp. 311–328.

Llanos-Pérez, JA, Betancourt-Mar, A, De Miguel, MP, Izquierdo-Kulich, E, Royuela-García, M, Tejera, E, & Nieto-Villar, JM 2015 "Phase transitions in tumor growth: II prostate cancer cell lines", *Physica A: Statistical Mechanics and its Applications*, vol. 426, pp. 88–92.

Llanos-Pérez, JA, Betancourt-Mar, JA, Cocho, G, Mansilla, R, & Nieto-Villar, JM 2016 "Phase transitions in tumor growth: III vascular and metastasis behavior", *Physica A: Statistical Mechanics and its Applications*, vol. 462, pp. 560–568.

Lloyd, S 2001 "Measures of complexity: a nonexhaustive list", *IEEE Control Systems Magazine*, vol. 21, no. 4, pp. 7–8.

Lucia, U 2014 "Entropy generation and cell growth with comments for a thermodynamic anticancer approach", *Physica A: Statistical Mechanics and its Applications*, vol. 406, pp. 107–118.

Lucia, U, Ponzetto, A, & Deisboeck, TS 2015 "A thermodynamic approach to the 'mitosis/apoptosis' ratio in cancer", *Physica A: Statistical Mechanics and its Applications*, vol. 436, pp. 246–255.

Luo, LF 2009 "Entropy production in a cell and reversal of entropy flow as an anticancer therapy", *Frontiers of Physics in China*, vol. 4, pp. 122–136.

Mansilla, R & Nieto-Villar, JM (eds.) 2017 *La Termodinámica de los sistemas complejos*. Mexico City: UNAM.

Marin, D, & Sabater, B 2017 "The cancer Warburg effect may be a testable example of the minimum entropy production rate principle", *Physical Biology*, vol. 14, no. 2, pp. 1478–3975.

Martin, RR, Montero, S, Silva, E, Bizzarri, M, Cocho, G, Mansilla, R, & Nieto-Villar, JM 2017 "Phase transitions in tumor growth: V what can be expected from cancer glycolytic oscillations?", *Physica A: Statistical Mechanics and its Applications*, vol. 486, pp. 762–771.

Martyushev, LM, & Seleznev, VD 2006 "Maximum entropy production principle in physics, chemistry and biology", *Physics Reports*, vol. 426, no. 1, pp. 1–45.

Mawhin, J 1996 "The early reception in France of the work of Poincaré and Lyapunov in the qualitative theory of differential equations", *Philosophia Scientiae*, vol. 1(4), pp. 119–133.

Mesa-Rodríguez, A., Gonzalez, A, Estevez-Rams, E, & Valdes-Sosa, PA, 2022 "Cancer segmentation by entropic analysis of ordered gene expression profiles", *Entropy*, vol. 24, no. 12, p. 1744.

Michaelian, K 2022 "Non-equilibrium thermodynamic foundations of the origin of life", *Foundations*, vol. 2, no. 1, pp. 308–337.

Miquel, J, Economos, AC, & Johnson Jr, JE 1984 "A systems analysis—thermodynamic view of cellular and organismic aging". In: Johnson, JE (ed.) *Aging and Cell Function.* , pp. 247–280, Boston, MA: Springer US.

Molnar, J, *et al.* 2005 "Thermodynamic aspects of cancer: Possible role of negative entropy in tumor growth, its relation to kinetic and genetic resistance", *Letters in Drug Design & Discovery*, vol. 2, no. 6, pp. 429–438.

Montemayor-Aldrete, JA, Márquez-Caballé, RF, del Castillo-Mussot, M, & Cruz-Peregrino, F 2020 "General thermodynamic efficiency loss and scaling behavior of eukaryotic organisms", *Biophysical Reviews and Letters*, vol. 15, no. 03, pp. 143–169.

Montero, S, Martin, R, Mansilla, R, Cocho, G, & Nieto-Villar, JM 2018 Parameters estimation in phase-space landscape reconstruction of cell fate: A systems biology approach. In: Bizzarri, M (ed.) *Systems Biology: Methods in Molecular Biology*, vol. 1702. New York: Humana Press. https://doi.org/10.1007/978-1-4939-7456-6_8.

Nicolis, G, & Nicolis, C 2007 *Foundations of Complex Systems. Nonlinear Dynamics. Statistical Physics. Information and Prediction*. Singapore: World Scientific Publishing Co. Pte. Ltd.

Nicolis, G 1972 "Fluctuations around nonequilibrium states in open nonlinear systems", *Journal of Statistical Physics*, vol. 6, pp. 195–222.

Nicolis, G, & Daems, D 1998 "Probabilistic and thermodynamic aspects of dynamical systems", *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 8(2), pp. 311–320.

Nicolis, G, & Prigogine, I 1977 *Self-Organization in Nonequilibrium systems*. New York: Wiley.

Nicolis, G, & De Decker, Y 2017 "Stochastic approach to irreversible thermodynamics", *Chaos*, vol. 27, no. 10, art. 104615. https://doi.org/10.1063/1.5001303

Nicolis, G, & Nicolis, C 2000 "What can we learn from thermodynamics on stochastic dynamics?" In: *AIP Conference Proceedings*, vol. 502, no. 1, pp. 333–348. Ambleside: American Institute of Physics.

Nieto-Villar, JM, Betancourt-Mar, J, Izquierdo-Kulich, E, & Tejera, E 2013 *Complejidad y Auto-organización en Patrones Naturales*. Editorial UH.

Nieto-Villar, JM, Quintana, R, & Rieumont, J 2003 "Entropy production rate as a Lyapunov function in chemical systems: Proof", *Physica Scripta*, vol. 68, no. 3, pp. 163–165.

Nieto-Villar, JM, García-Fernández, JM, & Rieumont-Briones, J 1995 "The rate of entropy production as an evolution criterion in chemical systems: I. Oscillating reactions", *Physica Scripta*, vol. 52, no. 1, pp. 30–32.

Nieto-Villar, JM, Izquierdo-Kulich, E, Quintana, R, & Rieumont, J 2013 "Una aproximación del criterio evolutivo de Prigogine a sistemas químicos", *Revista Mexicana de Física*, vol. 59, no. 6, pp. 527–529.

Nieto-Villar, JM, & Mansilla, R 2022 "Longevity, aging and cancer: Thermodynamics and complexity", *Foundations*, vol. 2, no. 3, pp. 664–680.

Nieto-Villar, JM; & Mansilla, R 2021 "Ferroptosis as a biological phase transition I: Avascular and vascular tumor growth", *European Journal of Biomedical and Pharmaceutical Sciences*, vol. 8, no. 12, pp. 63–70.

Nieto-Villar, JM, Rieumont, J, & Mansilla, R 2022 "The entropy production rate a bridge between thermodynamics and chemical kinetics", *Revista Mexicana de Física E*, vol. 19, no. (1 Jan-Jun), art. 010212-1.

Nieto-Villar, JM, Rieumont, J, Quintana, R, & Miquel, J 2003 "Thermodynamic approach to the aging process of biological systems", *Revista CENIC Ciencias Químicas*, vol. 34, no. 3, pp. 149–157.

Nieto-Villar, JM, & Velarde, MG 2001 "Chaos and hyperchaos in a model of the Belousov-Zhabotinsky reaction in a batch reactor", *Journal of Non-Equilibrium Thermodynamics*, vol. 25, no. 3–4, pp. 269–278.

Norton, L 2005 "Conceptual and practical implications of breast tissue geometry: Toward a more effective, less toxic therapy", *The oncologist*, vol. 10, no. 6, pp. 370–381.

Onsager, L 1931 "Reciprocal relations in irreversible processes I", *Physical Review*, vol. 37, pp. 405–426.

Onsager, L 1931 "Reciprocal relations in irreversible processes. II", *Physical Review*, vol. 38, pp. 2265–2279.

Prigogine, I 1947 *Etude thermodynamique des phenomenes irreversibles*. Thèses d'agrégation de l'enseignement supérieur de l'Université Libre de Bruxelles. Paris: Dunod.

Prigogine, I 1961 *Introduction to Thermodynamics of Irreversible Processes*. New York: Wiley.

Prigogine, I 1978 "Time, structure, and fluctuations", *Science*, vol. 201, no. 4358, pp. 777–785.

Rieumont-Briones, J, Nieto-Villar, JM, & García, JM 1997 "The rate of entropy production as a mean to determine the most important reaction steps in Belousov-Zhabotinsky reaction", *Anales Química*, vol. 93, pp. 147–152.

Rényi, A 1960 "On measures of information and entropy", In: *Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability*, vol. 1, pp. 547–561. Berkeley and Los Angeles: University of California Press.

Schuster, HG, & Just, W 2006 *Deterministic Chaos: An Introduction*. Weinheim: Wiley-VCH.

Strogatz, SH 2000 *Nonlinear Dynamics and Chaos*. Boulder: Westview Press.

Triana, L, Cocho, G, Mansilla, R, & Nieto-Villar, JM 2018 "Entropy production as a physical pacemaker of lifespan in mole-rats", *International Journal of Aging Research*, vol. 1, no. 5, pp. 1–9.

Turányi, T. 1990, "Sensitivity analysis of complex kinetic systems. Tools and applications", *Journal of Mathematical Chemistry*, vol. 5, no. 3, pp. 203–248.

Turanyi, T, Gyorgyi, L, & Field, RJ 1993 "Analysis and simplification of the GTF model of the Belousov-Zhabotinskii reaction", *The Journal of Physical Chemistry*, vol. 97, no. 9, pp. 1931–1941.

Varma, A, Morbidelli, M, & Wu, H 2005 *Parametric Sensitivity in Chemical Systems*. Cambridge: Cambridge University Press.

Organisms

ISSN: 2532-5876