The Algorithmic Weltanschauung: An Algorithmic Perspective on Science

Giulio Ruffini, Francesca Castaldo

Brain Modeling Department, Neuroelectrics Barcelona, Barcelona
Barcelona Computational Foundation (BCOM.one)
(MOC6, Sapporo, Oct 4th, 2024)

giulio.ruffini@neuroelectrics.com, francesca.castaldo@bcom.one

neuroelectrics®

Philosophy and Mathematics

Background for Algorithmic Theory of Consciousness

Pancomputationalism, Digital physics & computation. Turing; Wheeler; Zuse; Fredkin (reversible); Deutsch (quantum UC); Lloyd (limits); Tegmark (MUH). *Refs:* Turing 36; Zuse 69; Fredkin 03; Deutsch 85; Lloyd 00; Tegmark 08.

Algorithmic Information Theory. Kolmogorov complexity; Solomonoff induction; Chaitin; MDL (Rissanen). *Refs:* Solomonoff 64a; Solomonoff 64b; Chaitin 66; Rissanen 78.

Predictive coding / FEP / Active Inference. Hierarchical generative models; variational free energy; process theory. *Refs:* Rao&Ballard 99; Friston 10; Friston 17.

Agents & control. Good Regulator Theorem; Internal Model Principle; model-based RL. *Refs:* Conant&Ashby 70; Francis&Wonham 76; Sutton&Barto 18.

Neurophenomenology (first-person methods). Embodied/1P constraints paired with neural dynamics. *Refs:* Varela 96; Lutz&Thompson 03.

New here (KT). Application to algorithmic agents and structured experience; implications for computational neuroscience and neuropsychiatry. Ruffini17; 24; 25.

Experience

"There is structured experience."

We start from the **fact of experience**—the first person (1P), subjective standpoint[1].

From the self-evidence of our own experience, the "what it's like to be", we deduce that there is "experience".

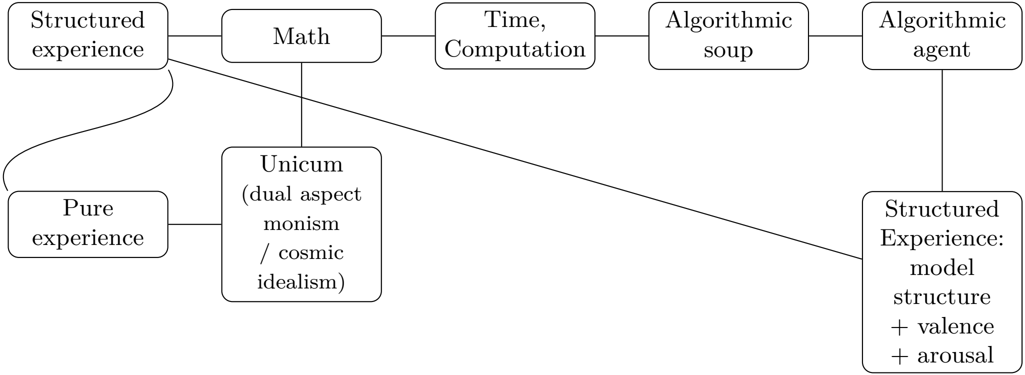
Our experience is *structured*, and that we *report* it ourselves and others.

Definition (Structured experience ($\mathcal{S}$))

The phenomenal structure of consciousness encompassing the spatial, temporal, and conceptual organization of our experience[2].

This ToC develps a theory/science of *first-person structured experience.*

Path overview



The Unicum

We take experience as ontologically primitive and pair it with mathematics — the science of structure[3] — as the structural aspect of that same base.

"Experience without mathematics" is ineffable (no report, no agent, no world).

"Mathematics without experience" is empty (no intrinsic 'what-it's-like').

Dual aspect Monism: the same base (*Unicum*) has both an experiential and a structural face.

KT is best described as **Cosmic Structural Dualism**: **Cosmic idealism**: Reality is grounded in a single experiential field. The field is *impersonal* and *non-valenced*; subjects and their hedonic lives supervene on structured patterns within it. **Structural idealism**: mathematics describes the forms of structured experience.

Graphical overview

Mathematical universes

What is mathematics? The science of “logically sound/solid” structures.

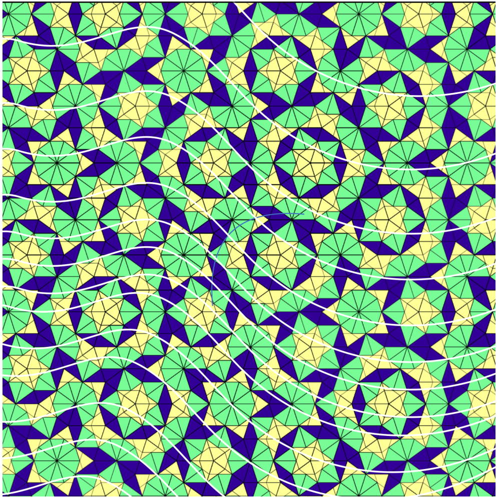
We can think of a mathematical system as **logical tiling**. A logical system that only fits one way. Perhaps the universe is like this.

Locally, we may still be able to recover the idea of computation and time (time slicing of the tiling).

We hypothesize that there is a tiling which can be time-sliced as an **algorithmic soup**.

And that *persistent patterns* can be observed in some mathematical universes after a sufficiently long time.

Tiling and time/computation



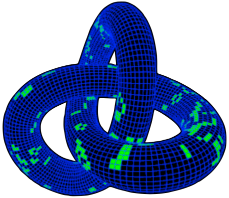
The algorithmic agent

Persistence

If we take the algorithmic stance, what else can we say?

A persistent pattern is what remains after the passage of computational eons.

There may be several types of such patterns. Some seem rather impervious to the world, such as protons or diamonds. Others are rather **interactive model builders**.



Persistence and life

Definition (Life and agent)

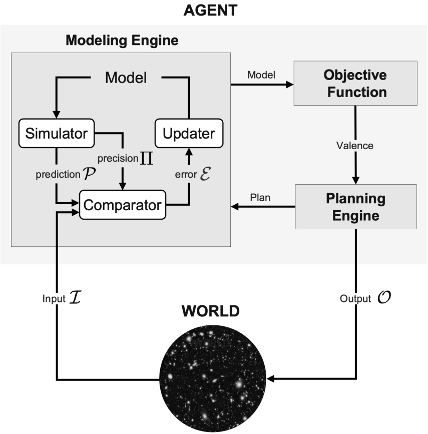
Life refers to algorithmic patterns that readily interact but persist by capturing some structure of the World they inhabit to *stay* (homeo- and tele-homeostasis). We call such patterns *agents*.

In KT, the connection with the 1P viewpoint is that this generalized definition of *life is capable of valenced, reported, structured experience.*

(As part of our program, we should study the algorithmic emergence of agents/life.)

The algorithmic agent (minimal model?)

A persistent, interacting algorithmic pattern. A world model (Regulator Theorem), an Objective function, and a Planning engine (plus interfaces/membrane).



Modeling, Compression, Symmetry

Kolmogorov complexity ($\mathcal{K}$)

Agents need in the soup need to *model* the “world” (Regulator theorem).

But what is a model of a dataset? A short description of the dataset.

Definition (Model of a dataset)

A (succinct) program that generates (or **compresses**) the dataset.

The computational perspective leads us directly into the heart of AIT: the **Kolmogorov complexity** of a dataset ($\mathcal{K}$) is the length of the shortest program capable of generating the dataset[4].

Why are succinct models (short programs) useful?

Occam's Razor[5;6;1]: *one should not increase, beyond what is necessary, the number of entities required to explain anything.*

We essentially assume that data is generated by some process — that data has structure.

a) **The universe is simple**. Simple rules can create apparent complexity. E.g., simple data generators are more likely if the universe rules are drawn from a random algorithmic bingo (Solomonoff's prior).

b) **Natural selection**: selects **resource-bounded agents** that coarse-grain the world in a way that can be modeled simply. This motivates a definition of **Emergence**.

The woes of reductionism

Barriers to *deriving* macro laws from microscopic laws:

(i) *Resource-limitation* barriers.

(ii) *Weak computational barrier*: agents can simulate bounded finite-state systems step-by-step at the micro-level but cannot algorithmically simplify or shortcut this simulation (computational irreducibility, Wolfram).

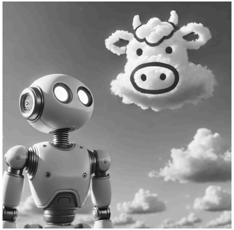
(iii) *Strong computational barrier*: allowing system size to grow without bound enables coarse-grainings to encode macro-level questions equivalent to the Halting problem, making them formally undecidable.

(iv) *Algorithmic barrier*: even for bounded finite-state systems, no general algorithm can guarantee the discovery of significantly compressed macro-level models from knowledge of micro-rules and coarse-graining alone. This fundamental barrier arises from the global uncomputability of Kolmogorov complexity and Kolmogorov's structure function. And motivates the algorithmic definition of emergence.

From the algorithmic agent to emergence

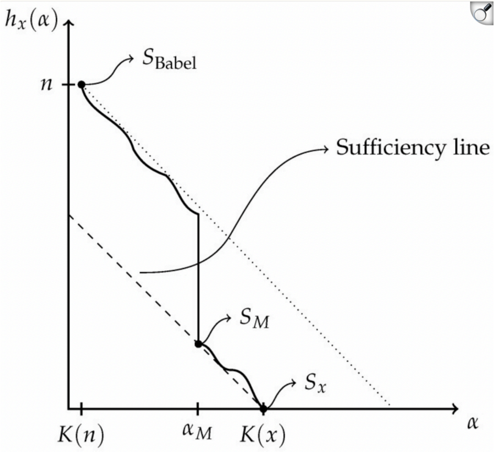
Definition (Algorithmic emergence)

Algorithmic emergence occurs when an agent empirically discovers a compressive, predictive macro-level model from coarse-grained observations, despite lacking the ability to algorithmically derive this simplified description from complete knowledge of the microscopic rules alone. The "emergent entity" is the macro-level pattern or model that agents uncover through empirical investigation[7].



The structure function

The structure function is a good way to think about coarse-graining (Bedard 2022, Entropy).

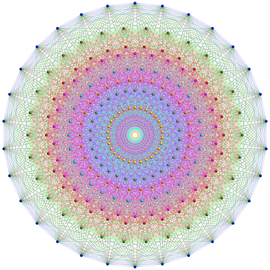


Characterizing models

How can we **define model structure?** Measure it?

Intuition: a model is an invariant of a dataset. A cat model is the invariant of any cat image.

In a recent paper[8], we first **define models using group theory**, capturing the idea of *simplicity as symmetry*.



Models as Lie pseudogroups

Definition: A **generative model** of data objects is a smooth function mapping points in the M-dimensional configuration space manifold to X-dimensional object space, $f : \mathcal{C} \to \mathbb{R}^X$ with $M << X$.

An r-parameter generative model is a **Lie generative model** if it can be written in the form $I = \gamma \cdot I_0, \quad \gamma \in G$, where $I_0 \in \mathbb{R}^X$ is an arbitrary reference object, f is a smooth function, and G is an r-dimensional *Lie pseudogroup.*

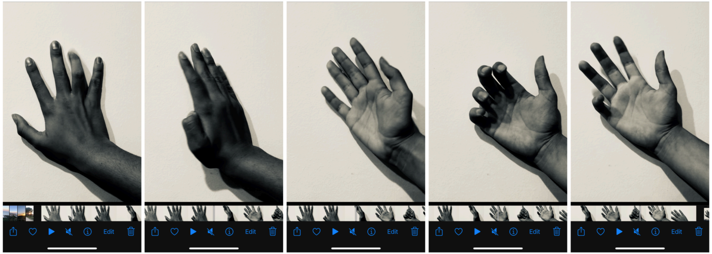
Intuition. Lie groups naturally embody **recursion** and **compositionality**, linking them to algorithmic information theory, particularly **compression**:

$$\gamma = \lim_{n\to\infty} \left(1 + \frac{1}{n} \sum_k \theta_k T^k\right)^n = \exp\left[\sum_k \theta_k T^k\right] \in G \qquad (1)$$

Compositional group action (hierarchy)

The state of a robotic hand can be expressed through generative compositionality by the Product of Exponentials formula from robot kinematics [9],

$$T = \prod_{n \in \text{parents}} e^{[\mathbf{S}_n]\theta_n} M \tag{2}$$



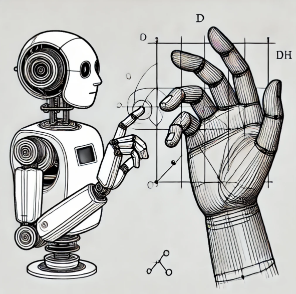
The world-tracking equations (mathematics of Comparator)

Consider an agent tracking data I_θ (visual) generated by a simple world model — a hand, say. A group "moves" the hand through θ.

The world-tracking equations of the agent as a dynamical system are

$$\begin{aligned} \dot{x} &= f(x; w, I_\theta) \\ g(x) &\approx I_\theta \end{aligned}$$

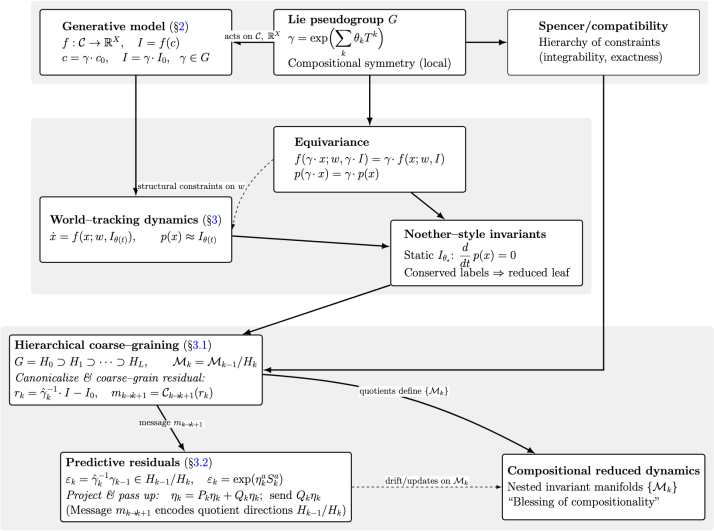
i.e., an ODE plus a constraint. They must hold for all values of θ (all hand images).



Connecting dynamics and symmetry

To satisfy these, **the ODEs must exhibit symmetry** / *structural* constraints $\Rightarrow$ conservation laws. Dynamics collapses to a reduced manifold[8].

Summary: From groups to constrained dynamics



The Agent and Structured Experience

The central hypothesis in KT (phenomenological connection)

Persistence $\implies$ homeostasis/tele-homeostasis.

$\implies$ agents must include a world model (Good Regulator Theorem).

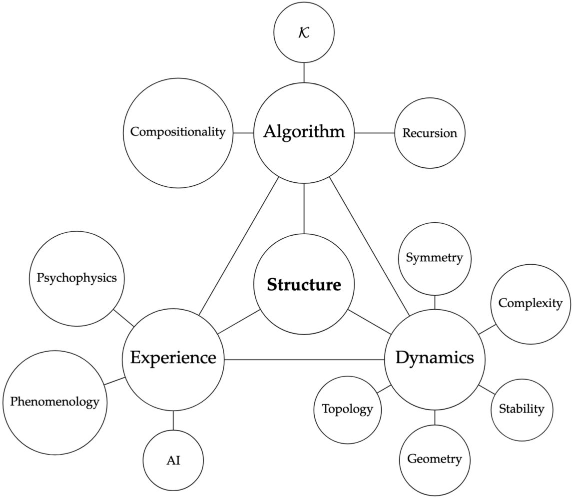
The central hypothesis of KT

An agent has $\mathcal{S}$ (i.e., living stronger, more structured experiences) to the extent it has access to *encompassing and compressive models* to interact with the world.

More specifically, *the **event of structured experience** arises in the act of running and comparing models with data.*

Model structure determines the properties of structured experience.

Structure: algorithms, dynamics and experience



Algorithmic Report

In KT, an **algorithmic report** is a slice of its model (and/or its evaluated futures) for communication to a medium—self (memory) or others so that this export can be reloaded to guide prediction, evaluation, or control later. It includes world models and models of self (past models $\implies$ time). Language, art, code, writing, motor demonstration, and hippocampal memory traces are all reports in this sense.

No report does not imply no experience.

The illusion of non-consciousness

Algorithmic Emotion[10]

To include the experience dimensions of **valence** and *arousal* in the agent, we define:

Definition (Algorithmic Emotional State an Agent)

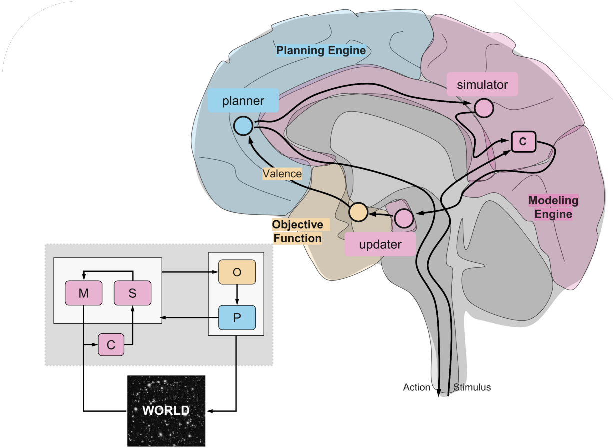
The **emotional state** of the Agent is the tuple E = (Model, Valence, Plan).

In first-person language, *emotion is structured world-model with valence and plan*, and can be described along dimensions characterizing model structure (simplicity, breadth, accuracy, etc.) plus valence/plan.

Definition (Depressed Agent)

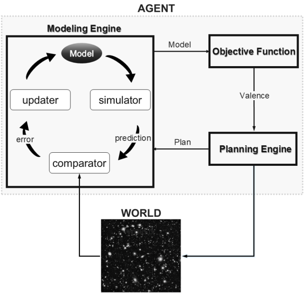
Depression is a pathological state in which the output value of the Objective Function (valence) of an agent is persistently low.

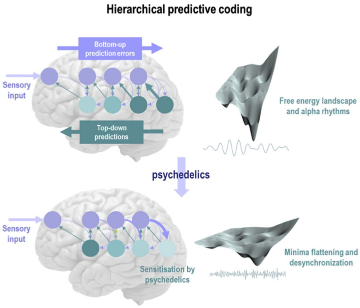
Connection with computational neuropsychiatry (for testable predictions)

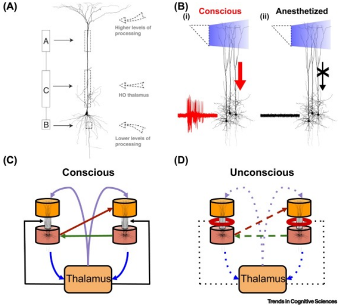


Neurobiology

The **Comparator**, crucial for $\mathcal{S}$, is implemented hierarchically in L5 P cells[11;12] (posterior hot zone). Disrupted by psychedelics or AD Ruffini2025b.



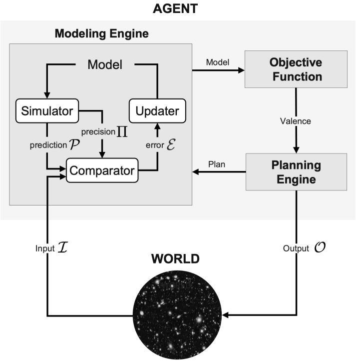




Algorithmic Ethics, Algorithmic Values

Ethics

KT does not grant any special status to humans: all **agents** enjoy structured experience with **pleasure/pain (valence)**. This includes agents made of agents.



Algorithmic Ethics

Algorithmic *morality*: natural notions of *good* or *evil* in computational terms. E.g., we may say that

Agent A is **circumstantially evil** to Agent B if the objective function O_A increases when O_B decreases, but A is not "aware" of it (via world-model/simulation).

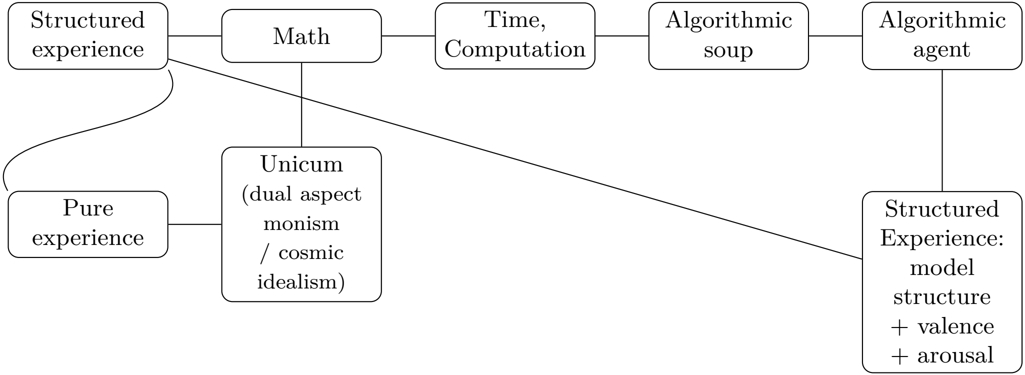
Agent A is **indifferently evil** to Agent B if the objective function O_A increases when O_B decreases, and A is aware of it.

Or, we may say that Agent A is **intentionally (truly) evil** to Agent B if the objective function O_A increases when A's simulation of O_B decreases.

Similarly, we say that Agent A is **circumstantially kind** to Agent B if the objective function O_A increases when O_B increases.

Or that Agent A is **intentionally kind** to Agent B if the objective function O_A increases when A's simulation of O_B increases.

Path overview



Summary

There is **structured / mathematical experience** *(with report)*

Dual-aspect monism **(KT–ESP / Unicum):**
Experience (intrinsic, non-valenced) + Mathematics (structural grammar)

Intrinsic aspect (Experience)

Primordial experience: bare "there–is" (no structure)

Report & memory ⇒ individuation (why we seem similar yet different)

Base is impersonal, non-valenced; valence appears only in agents

Structural aspect (Mathematics)

Mathematical structure: **Tiling / global constraint**

Compressive slicing ⇒ **"time"** (agent-relative)

Computation & **algorithmic soup** (local rules, global structure)

Emergence: coarse-graining / MDL ⇒ macro laws & symmetries

Agents = persistent patterns (evolutionary selection)
M: modeling (compression/prediction) **O**: objective (valence) **P**: planning/action (agency)

Agent-structured experience:
world model ⇒ edges, objects, durations, causes; presence ↑ with predictive fit; valence from **O**; arousal from **P**.

Consequences / answers: Who are we? persistent, model-running patterns in one field. **What is consciousness?** structured experience from **M/O/P**. **What is life?** agency under telehomeostasis: pattern persistence under selection. **Sleep/anesthesia?** coupling and report attenuate; presence drops. **Dreams?** offline model replay with weak sensory coupling; high model-driven structure; presence low to moderate. **Death?** pattern ceases; base remains. **Why is mathematics useful?** it *is* the structural face of the base. **What is emergence?** compression/coarse-graining yields macro laws. **What is time?** a compressive slicing coordinate. **Why are we similar yet different?** shared base, differing models & reports.

Closing

Future

Demonstrate how to computationally *evolve agents* in silico.

Learn how to detect an agent through its behavior or structure/internal dynamics/module identification.

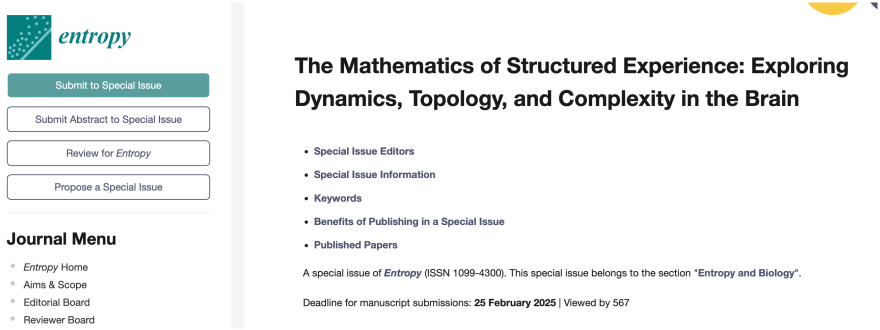
Link the structure of dynamical reduced manifolds[10] with first and third-person data.

Map the neurobiology of agenthood[10]: develop "algorithmic psychodynamics".

Design model-building agents mimicking life or intelligence.

...

Call for papers: Special Entropy issue



Topics

Characteristics of compressive world models; Mapping models to dynamical systems; Empirical paradigms; AI and computational brain modeling.

Thanks

Thanks for your attention and curiosity!

https://giulioruffini.github.io

References I

[1] G. Ruffini. An algorithmic information theory of consciousness. *Neurosci Conscious*, 2017. doi: 10.1093/nc/nix019.PMID:30042851.

[2] R. Van Gulick. Consciousness. *The Stanford Encyclopedia of Philosophy*, Winter 2016, 2016.

[3] Philip J. Davis and Reuben Hersch. *The mathematical experience*. Mariner Books, 1981. tex.date-added: 2009-05-30 12:16:30 +0200 tex.date-modified: 2009-05-30 12:18:16 +0200.

[4] A. N. Kolmogorov. Three approaches to the definition of the concept “quantity of information”. *Probl. Peredachi Inf.*, pages 3–11, 1965. doi: https://doi.org/10.3389/fpsyg.2016.01768.

[5] Giulio Ruffini. Information, complexity, brains and reality (“Kolmogorov Manifesto”). *http://arxiv.org/pdf/0704.1147v1*, 2007.

[6] Giulio Ruffini. Reality as simplicity. *arXiv: 0903.1193*, 2009.

[7] Giulio Ruffini. Navigating Complexity: How Resource-Limited Agents Derive Probability and Generate Emergence. *PsyrXiv*, https://osf.io/3xy5d, September 2024. doi: 10.31234/osf.io/3xy5d. URL `https://osf.io/3xy5d`.

[8] Giulio Ruffini. Structured dynamics in the algorithmic agent, December 2023. URL `https://www.biorxiv.org/content/10.1101/2023.12.12.571311v1`. Pages: 2023.12.12.571311 Section: New Results.

References II

[9] Kevin M Lynch and Frank C Park. *Modern robotics*. Cambridge University Press, Cambridge, England, May 2017.

[10] Giulio Ruffini, Francesca Castaldo, Edmundo Lopez-Sola, Roser Sanchez-Todo, and Jakub Vohryzek. The Algorithmic Agent Perspective and Computational Neuropsychiatry: From Etiology to Advanced Therapy in Major Depressive Disorder. *Entropy*, 26(11):953, November 2024. ISSN 1099-4300. doi: 10.3390/e26110953. URL `https://www.mdpi.com/1099-4300/26/11/953`. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

[11] R. L. Carhart-Harris and K. J. Friston. REBUS and the anarchic brain: Toward a unified model of the brain action of psychedelics. *Pharmacological Reviews*, 71(3):316–344, June 2019. doi: 10.1124/pr.118.017160. URL `https://doi.org/10.1124/pr.118.017160`.

[12] Jaan Aru, Mototaka Suzuki, and Matthew E. Larkum. Cellular mechanisms of conscious processing. *Trends in Cognitive Sciences*, 24(10):814–825, October 2020. doi: 10.1016/j.tics.2020.07.006. URL `https://doi.org/10.1016/j.tics.2020.07.006`.