Towards an algorithmic information theory of consciousness (KT)

Giulio Ruffini, Edmundo Lopez-Sola, Roser Sanchez-Todo

Brain Modeling Department
Neuroelectrics Barcelona, Barcelona

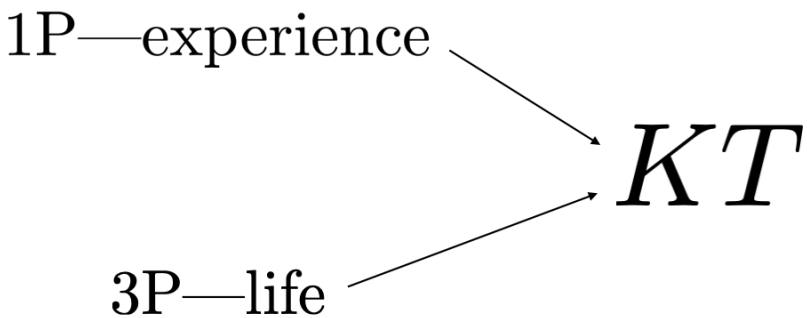
giulio.ruffini@neuroelectrics.com

April 19, 2022

Overview

Motivation

Two routes to KT



The subjective route to KT: Experience (1P)

We start from the fact of *experience*—the first person (1P), subjective standpoint [Ruf17].

Meditation, psychedelics, religious experience (e.g., Buddhism) suggest that experience can be pure/primordial, free from mental constructs such as the ego.

From the self-evidence of our own experience, the "what it's like to be", we assume that primordial experience does not allow for or require prior causes.

Warning: We assume *there exists experience.*

KT does not address the hard problem of consciousness.

Structured experience ($\mathcal{S}$)

We aim to build a theory around the notion of *structured experience*—where *mathematics* and experience meet.

Mathematics: The science of structure, order, and relation [Gra10].

We observe that experience is *structured*: at least during wakefulness, there is a spatial, temporal, and conceptual organization of our 1P experience of the world (inc. *self*).

Definition (*structured experience* ($\mathcal{S}$))

The phenomenal structure of consciousness encompassing both sensory qualia and the spatial, temporal, and conceptual organization of our experience [Van16].

Scientific strategy for the study of $\mathcal{S}$

This definition of $\mathcal{S}$ can be empirically explored in reporting humans, and we aim to characterize it with methods applicable to a broad range of systems.

The strategy will be to quantify the structure of experience from **1P reports** in humans and attempt to associate it with 3P data (e.g., EEG, fMRI or behavior) using mechanistic insights derived from neuroscience and mathematics.

With this knowledge at hand, we can then study (3P) other systems (non-reporting humans, other living species or artificial agents) and provide an educated guess about the agent's $\mathcal{S}$.

The objective route to KT (3P): persistence and life

We can also start by attempting to define what *life* is.

What remains after the passage of eons must rightfully be called a *persistent pattern*.

There may be several types of such patterns. Some seem rather impervious to the world, such as protons.

Definition (Life)

Patterns that readily interact but persist by partly capturing structure in the world they inhabit to stay or replicate (homeo- and meta-homeostasis).

The connection with the first viewpoint is that, in KT, this generalized definition of *life is what is capable of* $\mathcal{S}$.

As part of our program, we should study the algorithmics of the emergence of life.

AIT and Kolmogorov complexity

Kolmogorov complexity ($\mathcal{K}$) I

We can think of agents as physical systems, and in turn, of physical systems as dynamical systems calculating (effectively computable) functions. This allows us to analyze agents from the standpoint of computation theory.

Warning: Computation is a mathematical concept (Turing machine)

The use of computational framework in KT should not be construed to mean that the brain is literally a physical von Neumann computer (such as a laptop).

A computational perspective leads us directly into algorithmic information theory (AIT) and its central concept:

Definition (Kolmogorov complexity of a dataset $\mathcal{K}$)

The length of the shortest program capable of generating the dataset [Kol65, CT06].

Kolmogorov complexity ($\mathcal{K}$) II

Data

['3.141592653589793238126441332783153851346693837510901895726',
'0183844990527447207158188099987323024033062286954749281803 20',
'7536741059927308898181899002839603178193811773431727203959358',
'2681413334524778875085102698279378964782750493870397771769 00',
'4052415221610820855312843745938676064382634796104288333732 40',
'8145257445815990368592973580473662870200472728373853305809 90',
'3219954525817261060491726903951980625736652628712969609052 98',
'7671009626536868727705897236681911022946747224333945197705 300',
'0551194650093595826117491241484296185084399045107683259827 42',
'2796752757654895253426715085241676350146131019255952023557 55',
'1025857702005465094718990466067932265266791062245718982808 00',
'1079018037270642344517933906555987110685595058411279280600 32',
'3371928865285029664919303221161587360568283466470175472487 99',
'5047881060144389117952834809580569970793136663404028823381 00',
'8859878642076502910551119205187427530885652009759019838144 37',
'9959543614228588756397294345051927549776778582253957517462 94',
'1062524121123966056730332965125414383488745839700585848528 01',
'7589679472248565381117609951275291347605491550540934846424 04',
'1446507660740795192052525066129611766269075058004912488392 16',
'0959876470338606019660536434636214015643677347934323381160 39',
'0976580751658357898726311389441335589100941109793653014816 82',
'5098139534461506842152558590766974227268103727888582250211 20',
'0915431100227729084163271858880408201234196220085730263522 49',
'6617415036846819539299612901084677056805139712488551953236 0',
'4226226703436236014338139502905372278408088211001260582911 94',
'0710443484622851371228598214309735550913149138931732985440 9',
'7640031876246093545087677648166725209463342004957077468323059',
'4035067204193117534783667673388479448653806180626862653194 97',
'3434402584778120303661390088673738248660363539288600905926 86',
'3144203920372574405560242234215280676562246532983297176209 22',
'0959331949526318173011107900898395940425054357988500745683 20',
'2087643303992130536656797476045645066572460155175056566941 884',
'9484861086981528759507365950197093503574535765470278825068',
'6472743448866089944433814954585492799023580238978785646415 88',
'8274199009398310215360665972538336186638618863539243111550 72',
'5465632597341871693092005048369162816989158228242909338067 93',
'4236346495476442304739861284858223431494392904021289244957 74',
'5968725303622465325518055916047926857762066107005081060125 00',
'1379937745532168307261807405369060855092774778605409585758 41',
'7126238760037876516321431973866779929212623501645513508352 38',
'8611813000990276977397495558044340651038911414431714386451 33',
'2369551974083304070317338997181759848124743853743856318798 66',
'9771169973976839394704624349579551136117614953533238626691 238',

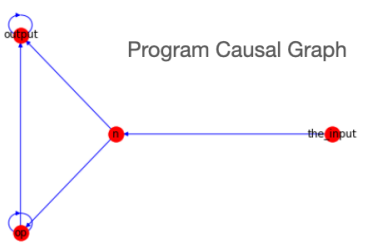
Program/model

```python
# https://www.wikihow.com/Write-a-Python-Program-to-Calculate-Pi
def nilakantha(the_input):

    variables=['the_input', 'n','op','output']
    dictvariables = { i : variables[i] for i in range(0, len(variables) ) }
    c = np.zeros((len(variables),len(variables)),dtype='int32') #from to

    output = Decimal(3.0)
    op = 1
    n = 2

    c[1,0] += 1
    for n in range(2, 2*the_input+1, 2):
            c[3,1 ] += 1; c[3,2 ] += 1; c[3,3] += 1
            output += 4/Decimal(n*(n+1)*(n+2)*op)
            c[2,2] += 1; c[2,1] += 1;
            op *= -1
    return output,c.transpose(),dictvariables
```



Mutual algorithmic information ($\mathcal{M}$)

With $\mathcal{K}$ at hand, we can define an algorithmic version of mutual information we will also need:

Definition (Mutual algorithmic information complexity $\mathcal{M}$)

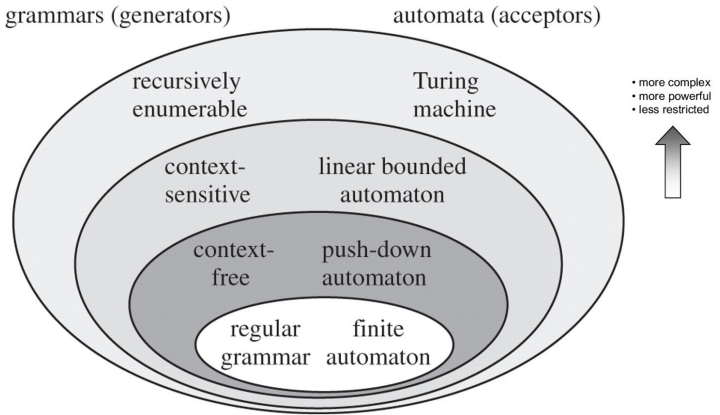
The *mutual algorithmic information* $\mathcal{M}(x:y)$ *between two strings* x *and* y*, is given by*

$$\mathcal{M}(x:y) = \mathcal{K}(x) + \mathcal{K}(y) - \mathcal{K}(x,y)$$

[LV97, GV04].

Hierarchy class [Fit14]

Not all programming languages are equal. Recurrence is needed for Turing completeness, for example.



Model I

The notion of **model** is central in KT and other theories of consciousness.

Definition (Model)

A model of a dataset is any program that generates the dataset.

Models may differ in two ways: they may implement different functions, or they may implement the same function in different ways. Both aspects matter here. We focus on those that implement the right functions succinctly.

Definition (Optimal model)

The optimal model of a dataset is the shortest program that generates (or, equivalently, compresses) the dataset.

Model II

An **optimal model needs to capture and exploit all the structure in the dataset**—and nothing else. In some sense, the *structure of the model can be described by the group of symmetries of the dataset* [Ruf16].

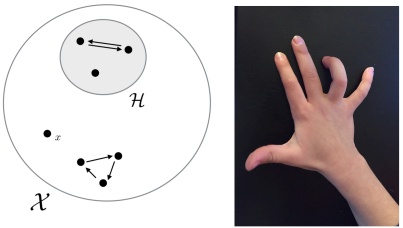
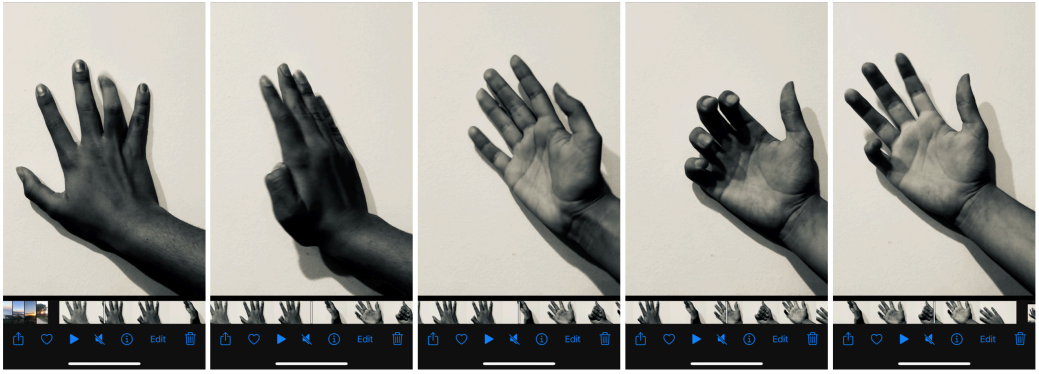


FIGURE 1. Left: Input (image) space $\mathcal{X}$, the subspace of hand images $\mathcal{H} = \{x \in \mathcal{X} | f(x) = 1\}$ and a example of automorphism (arrows) leaving the class one set $\mathcal{H}$ invariant. Right: a sample element $x \in \mathcal{H} \subset \mathcal{X}$.

Suppose we are given a stack of images of a hand, e.g., from the frames of a movie of a moving hand created using a generating function, $y = f(\theta)$, where y is the image in a frame and θ a parametrization of the hand image and view.

Model III



The structure of the dataset is encapsulated by the minimal program that encodes the function $y = f(x)$—the *invariant* object.

Why are "good models" good?

The rationale for the importance of compressive (succinct) models is discussed in detail in [Ruf07, Ruf09, Ruf17]. Rephrase of the principle of Occam's Razor: *one should not increase, beyond what is necessary, the number of entities required to explain anything.* Ok, but why?

The universe appears to be simple. Simple rules can create apparent complexity.

Simple data generators are more likely if the universe rules are drawn from a random algorithmic bingo (Solomonoff's prior).

Simple models are unbiased and generalize better (Occam, Laplace, Jaynes).

They are easier to construct, store, and reuse for model-building.

The central hypothesis

The central hypothesis in KT

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* (*good models*) to interact with the world.

More specifically, *the event of structured experience arises from the act of successfully comparing good models with data.*

Program structure determines the properties of structured experience.

The central hypothesis in KT

Compressive model refers to a simple, succinct program—the optimal model defined above is best possible scenario. Compressive models are special: to be short and accurate they need to capture structure in the data—low $\mathcal{K}$.

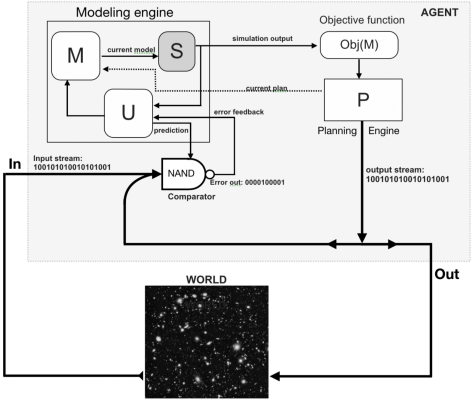
Encompassing refers to the amount of data from the world successfully matched, and hence explanatory potential. Eq., high $\mathcal{M}$.

Models are constructed from information generated by the sensorimotor system as the agent interacts with the external world: should account for data generated by the external world and by the agent itself—i.e., include a *self-model*.

The algorithmic agent

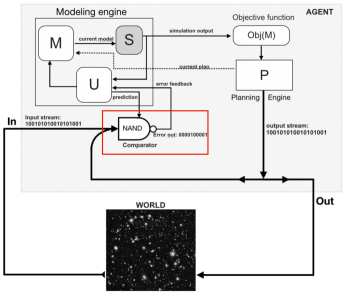
The algorithmic agent

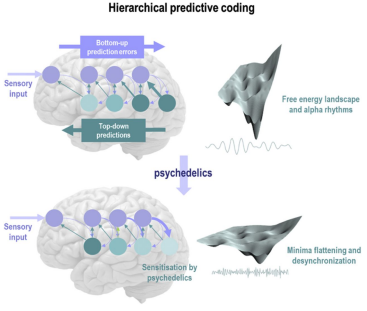
Minimal set of elements needed for a homeostatic algorithmic system. To be connected with (neuro)biology.

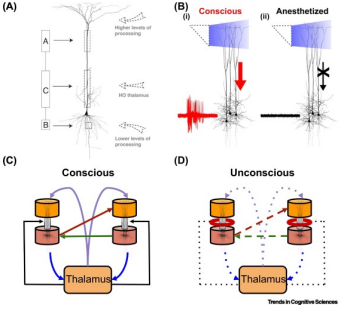


The algorithmic event of $\mathcal{S}$

Theories of cortical processing emphasize the separation of forward and backward information flow in the cortex also mirrored at the level of single cortical pyramidal cells [CHF19, ASL20]. **Comparator** implemented hierarchically in L5 P cells.







Model-building I: life

How do agents build models? In addressing this, we are led to connecting the concepts of life, intelligence, and $\mathcal{S}$.

Both life and intelligence represent processes to construct simple models for the persistence of algorithmic information-preserving systems across time.

Starting from *resilient building blocks (static persistence)*, from a computational perspective *life* is an algorithmic process: program building carried not solely by the individual agent, but by the transgenerational agent through evolution for *meta-homeostasis* (preservation of kind) (v. also [WD13, Cha12]).

Model-building II: intelligence

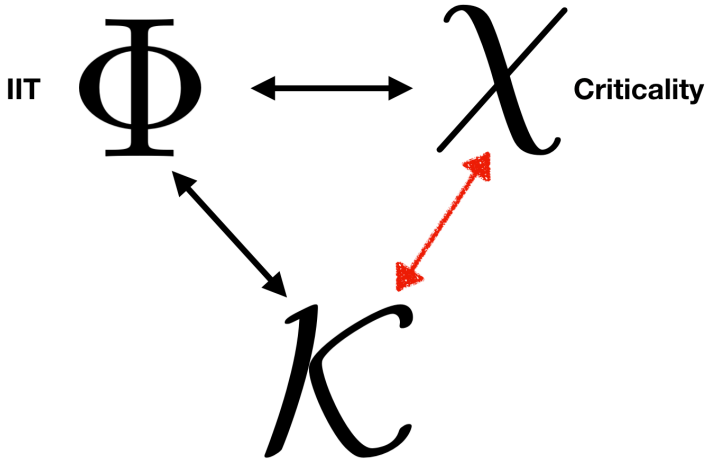
Evolutive pressure gives rise to the next leap, *intelligence*: agents that, starting from their static model (DNA in life) build higher-level compressive models of the world within their lifetime, e.g., using brains.

Importantly, KT holds that both static-model and active-modeling agents enjoy structured experience, only that their level of structure is possibly different.

[What comes after life and intelligence?]

Computation and dynamics

The criticality view I



The criticality view II

Computation in Nature is carried out by dynamical systems with very large degrees of freedom.

Brains operate close to such critical boundaries consistent with the notion of self-organized criticality (SOC) [BTW88, Chi04, CLZB17, CH18, DPS$^+$21].

Altered state of consciousness appear to move the brain away or towards criticality [CHF19, Rea22].

KT and criticality theory: Algorithmic agents (dynamical systems instantiating compressive models of data that exhibits regularities/symmetries) must have special properties. What are they?

The criticality view III

Recall the movie of a moving hand in empty space, $y(t) = f(\theta(t))$.

Although y may be embedded in a very high dimensional space, its dimension is actually very small if the set of parameters θ controlling the hand function is small.

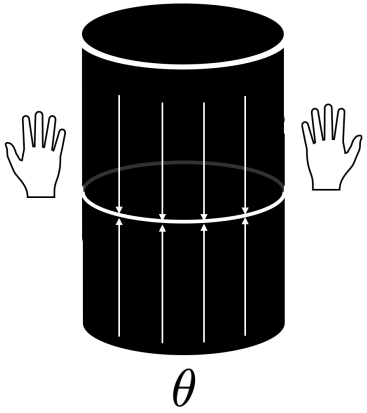
The state of a dynamical system generating frames of the moving hand, regardless of how large its natural space is (e.g., large number of neurons) must also lie in a low dimensional subspace, a **reduced manifold**. How?

Criticality: near criticality ($\mathrm{Re}[\lambda] \sim 0$) the dynamics of complex systems collapse to low dimensional manifolds [Jir20, JS22]—constrained dynamics.

Symmetry: in a Hamiltonian dynamical system where $g = y(t) - f(\theta(t)) = 0$ (the constraint), Noether's theorem states that H is invariant under the group of symmetries generated by g [Dir01, JS98]).

The criticality view IV—the center manifold

Trajectory of representation of hand in reduced manifold.



The criticality view V

Structure/symmetry in data, the collapse of dynamics to low-dimensional spaces, criticality (maximal information flow, power laws, long time scales and enhanced susceptibility), $\mathcal{K}$ and $\mathcal{S}$ are thus deeply connected.

The manifold structure of the reduced dynamics together with $\mathcal{M}$ provide, respectively, metrics on the simplicity of the models and the amount of algorithmic information captured.

The agent-world-lock loop—which tracks world data—keeps dynamics on the reduced manifold ($\mathcal{S}$!). Psychedelics, meditation, sensory deprivation, neuropsychiatric disorders, may "lift" up the enslaved dynamics.

Neurophenomenology of KT

Altered states of $\mathcal{S}$

Neurophenomenology defines a methodological strategy for integrating phenomenological and neurobiological accounts: 1P (phenomenology—subjective) and 3P (physiology, behavior—objective) data [Var96].

Altered states of consciousness such as psychedelics or meditation offer an interesting context to study the effect of perturbing the mechanisms of $\mathcal{S}$.

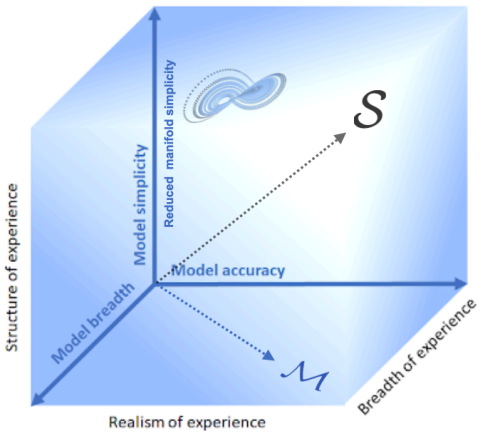
Meditative states are associated with the global dissolution of the embodied self [MCHR$^+$18] and can serve as unique models for investigation of self-dissolution (disengagement of self-models).

As an objective measure of structured experience, we can analyze descriptive narratives in speech form through state-of-the-art computational analysis (NLP) to establish metrics on text structure such as semantic coherence and speech disorganization index [SPC$^+$21, Tag22, MCR17].

The dimensions of structured experience

The structure of the reduced manifold represents the structure of experience, while model accuracy and breath ($\mathcal{M}$) map into the realism and breath of experience.

This is the map of the three dimensions of $\mathcal{S}$ into mathematical concepts.



Closing

Philosophy

KT is most naturally viewed in the context of panpsychism ('mind is everywhere', see Strawson, [Gof19]), a somewhat controversial version of the philosophy of consciousness.

Idealism is perhaps more rigorous philosophical background (consciousness as the fundamental entity, 'mind is everything').

Although not necessary to explore the scientific implications of the theory, the adoption of these can itself be motivated by simplicity and consistency criteria [Sym22].

Ethics

KT does not grant any special status to humans: all systems that capture structure from the world have structured experience.

Pleasure/pain associated with **objective function** $O \rightarrow$ morality: natural notions of *good* or *evil* in computational terms.

E.g., we may say that Agent's A is *evil* to Agent B if the objective function O_A increases when O_B decreases, that is $O_A(O_B)$ is decreasing or $O'_A(O_B) < 0$ (and viceversa for *good*).

Synergistic behavior emerges when agents are good to each other, while mutually-destructive behavior takes place in the complementary case.

Future

Much work remains to be done! Will KT provide a unification framework for the different approaches to consciousness? IIT, GWT, FEP, DIT all seem to fit. More work left to do in neurobiology of agenthood.

Can we computationally evolve agents? Are persistent patterns unavoidable (KT conjecture) if we wait long enough? Are there types other than static (proton), life and intelligence?

How can we discover the structure of reduced dynamics from physiological (system state) data?

Can we design better neurophenomenological methods to study $\mathcal{S}$?

Can we design model-building agents mimicking life or intelligence? Is AI the next evolutionary model-building leap? Brain-to-brain communication?

Thanks

Thanks for your attention and curiosity!

giulio.ruffini@neuroelectrics.com, @ruffini (Twitter)

Slides available at
https://github.com/giulioruffini/Ruffini-KT-Tucson-presentation-April-19-2022
Preprint is on the way.

References I

[ASL20] Jaan Aru, Mototaka Suzuki, and Matthew E. Larkum. Cellular mechanisms of conscious processing. *Trends in Cognitive Sciences*, 24(10):814–825, October 2020.

[BTW88] Per Bak, Chao Tang, and Kurt Wiesenfeld. Self-organized criticality. *Physical Review A*, 38(1):364–374, July 1988.

[CH18] Robin L. Carhart-Harris. The entropic brain - revisited. *Neuropharmacology*, 2018.

[Cha12] Gregory J Chaitin. *Proving Darwin*. Pantheon, New York, NY, May 2012.

[CHF19] 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.

References II

[Chi04] Dante R. Chialvo. Critical brain networks. *Physica A*, 340:756–765, 2004.

[CLZB17] Luca Cocchi, Leonardo L.Gollo, Andrew Zalesky, and Michael Breakspearac. Criticality in the brain: A synthesis of neurobiology, models and cognition. *Progress in Neurobiology*, 2017.

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. John Wiley & sons, 2 edition, 2006.

[Dir01] Paul Dirac. *Lectures on Quantum Mechanics*. Dover Books on Physics. Dover Publications, Mineola, NY, March 2001.

[DPS+21] Gustavo Deco, Yonatan Sanz Perl, Jacobo D. Sitt, Enzo Tagliazucchi, and Morten L. Kringelbach. Deep learning the arrow of time in brain activity: characterising brain-environment behavioural interactions in health and disease. July 2021.

References III

[Fit14] W. Tecumseh Fitch. Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews*, 11(3):329–364, September 2014.

[Gof19] Philip Goff. *Galileo's Error — Foundations for a New Science of Consciousness*. Penguin Random House UK, 2019.

[Gra10] John J Gray. Mathematics, 2010.

[GV04] Peter Grunwald and Paul Vitanyi. Shannon information and kolmogorov complexity. *arXiv:cs/0410002*, 2004.

[Jir20] Viktor Jirsa. Structured flows on manifolds as guiding concepts in brain science. In *Selbstorganisation – ein Paradigma für die Humanwissenschaften*, pages 89–102. Springer Fachmedien Wiesbaden, 2020.

References IV

[JS98] Jorge V Jose and Eugene J Saletan. *Classical dynamics*. Cambridge University Press, Cambridge, England, August 1998.

[JS22] Viktor Jirsa and Hiba Sheheitli. Entropy, free energy, symmetry and dynamics in the brain. *Journal of Physics: Complexity*, 3(1):015007, February 2022.

[Kol65] A. N. Kolmogorov. Three approaches to the definition of the concept "quantity of information". *Probl. Peredachi Inf.*, pages 3–11, 1965.

[LV97] Ming Li and Paul Vitanyi. *An introduction to Kolmogorov Complexity and its applications*. Springer, 1997.

[MCHR+18] Raphaël Millière, Robin L. Carhart-Harris, Leor Roseman, Fynn-Mathis Trautwein, and Aviva Berkovich-Ohana. Psychedelics, meditation, and self-consciousness. *Frontiers in Psychology*, 9, September 2018.

References V

[MCR17] Natália B Mota, Mauro Copelli, and Sidarta Ribeiro. Thought disorder measured as random speech structure classifies negative symptoms and schizophrenia diagnosis 6 months in advance. *npj Schizophrenia*, 3(1):1–10, 2017.

[Rea22] G. Ruffini and et al. Lsd-induced increase of ising temperature and algorithmic complexity of brain dynamics. *BioRXiv*, 2022.

[Ruf07] Giulio Ruffini. Information, complexity, brains and reality ("Kolmogorov Manifesto"). *http://arxiv.org/pdf/0704.1147v1*, 2007.

[Ruf09] Giulio Ruffini. Reality as simplicity. *arXiv: 0903.1193*, 2009.

[Ruf16] G Ruffini. Models, networks and algorithmic complexity. *Starlab Technical Note - arXiv:1612.05627*, TN00339(DOI: 10.13140/RG.2.2.19510.50249), December 2016.

References VI

[Ruf17] G. Ruffini. An algorithmic information theory of consciousness. *Neurosci Conscious*, 2017.

[SPC+21] Camila Sanz, Carla Pallavicini, Facundo Carrillo, Federico Zamberlan, Mariano Sigman, Natalia Mota, Mauro Copelli, Sidarta Ribeiro, David Nutt, Robin Carhart-Harris, et al. The entropic tongue: disorganization of natural language under lsd. *Consciousness and Cognition*, 87:103070, 2021.

[Sym22] Jack Symes. *Philosophers on consciousness*. Bloomsbury Academic, London, England, February 2022.

[Tag22] Enzo Tagliazucchi. Language as a window into the altered state of consciousness elicited by psychedelic drugs. *Frontiers in Pharmacology*, page 900, 2022.

References VII

[Van16] R. Van Gulick. Consciousness. *The Stanford Encyclopedia of Philosophy*, Winter 2016, 2016.

[Var96] F.J. Varela. Neurophenomenology: a methodological remedy for the hard problem. *Journal of Consciousness Studies*, 3(4):330–349, 1996.

[WD13] Sara Imari Walker and Paul C. W. Davies. The algorithmic origins of life. *Journal of The Royal Society Interface*, 10(79):20120869, February 2013.