mathematics

MDPI

Article

Is Every Cognitive Phenomenon Computable? †

Fernando Rodriguez-Vergara [1,2,*] and Phil Husbands [1]

[1] AI Research Group, University of Sussex, Falmer, Brighton BN1 9RH, UK; p.husbands@sussex.ac.uk
[2] Centre for Consciousness Science, University of Sussex, Falmer, Brighton BN1 9RH, UK
* Correspondence: f.rodriguez-vergara@sussex.ac.uk
† This paper is an extended version of our paper published in Proceedings of the 2025 Conference on Artificial Life: Cyphers of Life, Kyoto, Japan, 6–10 October 2025.

Abstract

According to the Church–Turing thesis, the limit of what is computable is bounded by Turing machines. Following from this, given that general computable functions formally describe the notion of recursive mechanisms, it is sometimes argued that every organismic process that specifies consistent cognitive responses should be both limited to Turing machine capabilities and amenable to formalization. There is, however, a deep intuitive conviction permeating contemporary cognitive science, according to which mental phenomena, such as consciousness and agency, cannot be explained by resorting to this kind of framework. In spite of some exceptions, the overall tacit assumption is that whatever the mind is, it exceeds the reach of what is described by notions of computability. This issue, namely the nature of the relation between cognition and computation, becomes particularly pertinent and increasingly more relevant as a possible source of better understanding the inner workings of the mind, as well as the limits of artificial implementations thereof. Moreover, although it is often overlooked or omitted so as to simplify our models, it will probably define, or so we argue, the direction of future research on artificial life, cognitive science, artificial intelligence, and related fields.

Keywords: computability; cognition; non-algorithmic cognition; protocognition; agency; autonomy; Turing machine

MSC: 00A30; 92B02; 93A02

1. Introduction

The use of computational devices has expanded greatly in recent decades, to the point that today, most people can easily conceive an algorithm as a sequence of simple instructions for carrying out some specific task (something that was probably much more obscure to past generations). We understand that an algorithm may not need to *solve a problem*; maybe we just want to print something, listen to music, or check the news in some social media application. This quick, intuitive understanding, however, often leads to follow-up questions. Is there a limit to what a computational device can do? Will it solve any mathematical problem as long as we provide the correct inputs? Will it become conscious with the right architecture and some specific program? Will it meaningfully *decide* what to do?

As usually happens with the most interesting things (such as life, the universe, or consciousness, to name a few), as soon as we try to put our finger on it, we quickly come to realize that beyond the intuitive, rather general ideas we use on a daily basis, to come up

Academic Editor: Sergio Rubin

Received: 26 September 2025
Revised: 26 January 2026
Accepted: 26 January 2026
Published: 2 February 2026

with something close to a definition is actually quite a hard task. As a matter of fact, this is not an exclusive problem of our daily conversations and, maybe without some component of irony, scientific endeavors usually consider the understanding of these interesting things to be their ultimate goal. All of these "simple" questions are somehow the entrance door, and the unreachable answer lies at the end of some never-ending hallway. Indeed, and regarding our topic at hand, there is yet no clear agreement on what the mind actually is and what its relation is to other concepts such as computation, intelligence, or cognition. On the contrary, our ideas about what can be conceived as a mind sometimes seem to become more obscure the more we understand about these related concepts individually.

For about half a century now, a biological approach to cognition has gradually but heavily influenced our understanding of cognitive phenomena, shifting our previous conception from a rather anthropocentric and strongly computationalist view to one where cognition seems to be deeply entangled with life [1–4]. This has led us to reconsider the meaning of fundamental notions such as intelligence, behavior, computation, and of course cognition and even life to some extent.

There has been, nonetheless, a relatively recent theoretical push in the "opposite" direction, which has invited us to reconsider some specific cognitive aspects in terms of a raw intelligence that could be implemented ubiquitously; that is, multiple realizable processes that can be implemented by living organisms in some form but that are not exclusive to them [5–10]. This pendular motion between computationally and biologically based approaches will serve as the framework for this paper. From our point of view, as we develop it through this paper, many of the old antagonist positions seem to be reconcilable these days, which drives this discussion toward a new dilemma, namely the hypothetical limits of what we can explain through our current formal models and what these really stand for [11].

The central underlying idea is that, insofar as a problem or a sequence of mechanisms may be well defined in algorithmic terms, there will be some syntactic machine capable of carrying it out. Hence, *if* natural systems can be conceived of as decision-making entities, then their decisions should be characterizable as series of locally logical events (even if incredibly complex) and therefore formalizable in terms of some abstract, analogous device representing their (environmental) inputs, internal states, and state transitions. This, however, seems to unavoidably end up removing something essential from the picture, something for which there is no clear scientific concept yet but that is so deeply rooted in our experience that we wish to trace a line that separates *us* from what is computable or, maybe better, from what computers will ever do [4,12–15].

This contradiction between computable-like mechanisms and the hypothetical and strong intuition that systems with a mind somehow can overcome such restrictions is the conundrum underpinning this paper. In particular, we discuss the feasibility of the idea that the Church–Turing thesis not only marks the limit of what is formally computable but physically realizable. In other words, whether every form of mechanistic intelligence is bounded by the capabilities of Turing machines regardless of the specific instantiation (e.g., cellular automata, deep, spiking, or any other form of neural networks, and even quantum computation, to name some well-known cases), and, more importantly, whether we deem our own cognitive faculties to be a form of mechanistic intelligence of this kind or not.

In this paper, we suggest that cognitive science has long treated the Church–Turing thesis not only as a limit on computation but rather as a tacit boundary on cognition itself. This assumption has led to a persistent search for traits that can be described through computable functions, hence reinforcing the idea that understanding what the mind is requires a particular kind of mathematical formalization. This methodological commitment risks mistaking formal describability for physical reality. In fact, life and cognition may

not wholly abide to these constraints, and a rigorous examination of non-algorithmic possibilities may be a necessary step toward further understanding.

2. Conceptual Foundations

Around the ninth century, in Persia, algebra originated, providing a tool for people to solve everyday problems (like dividing cattle or land among interested parties) without the need to fully understand the mathematical basis of the solution but just the bare facts (e.g., the number of animals and people or the shape and size of some terrain) [16]. In modern words, we could say that people had to provide the inputs to a symbolic abstraction that, operationalized by a human following a sequence of steps (an algorithm), led to an output, a mechanistic manner to solve problems by any person capable of following such instructions.

With this in mind, perhaps a good way to intuitively understand computation is as a second-order automation or as an automatized algebra, whereby the human *operator* is physically replaced by some kind of machine that follows the symbolic instructions from inputs to outputs, a mechanistic way to solve problems with any machine capable of physically instantiating the symbolic objects and following the instructions.

In fact, theory of computation is a specific branch of mathematics that refers to *recursive logical steps* by which problems can be effectively solved, i.e., algorithmically. The underlying premise is that these problems can be resolved mechanically (in the sense of automatically); that is, they can be solved without requiring mathematical knowledge about the logical steps involved [17,18]. It underpins computation as instantiated on general-purpose digital computers.

It is important to consider that at the time of the introduction of modern ideas about computation, the term computer was used to refer to people who had the job of following instructions from tables to perform engineering calculations through prespecified procedures [18,19]. In this sense, given that human workers performed the arithmetic operations on their own, we could say that instruction tables had already externalized the algebra (what we today often think of as a *program* but not yet the data, nor the *machine*).

In that context, the interpretation of what an algorithm was, was based on the empirical evidence that many elements could be calculated following an *effective procedure* (a sequence of logical steps). This intuitive notion was replaced by a formal one, provided by the works of Church [20,21] and Turing [17]. Considering that computers at that time were humans, the idea of an algorithm as formalized by Turing [17], namely as a well-defined machine that could essentially perform the same job as a human computer with pen and paper, was most appealing. This gives us crucial insight into the link between the concept of computation and the human mind, namely the implicit view that machines, under the correct specifications, may be capable of performing any task, including the ones that we assume to be hallmarks of (human) intelligence, such as logic and language (as famously hinted at by Turing [22]).

2.1. What Do We Mean by Computational?

Nowadays, most of us have a good intuitive understanding of what computation is. Nevertheless, there are some subtle distinctions that may be of importance. Here we briefly present some simple mathematical fundamentals that may be useful as context.

A given set A is said to be computable if and only if its indicator function is computable. In simple words, this means that A is the subset of another set B and that there is a function f (the indicator function of A) that maps every element of A to one and any other element from B to zero:

$$1_A : B \mapsto \{0, 1\}$$

where the term 1_A is the common notation for the indicator function of the subset of a set (A and B here, respectively). Hence, given any element x taken from B, we can represent the indicator function for A a bit more formally with

$$1_A(x) = \begin{cases} 1 & if \quad x \in A \\ 0 & if \quad x \notin A \end{cases}$$

Usually, in the context of Turing machines, this is illustrated by making reference to a subset of the natural numbers so that we could say that a subset A of the set B is computable if there is a function f that, following a finite sequence of logical steps and given a number x as input, correctly decides whether x belongs to A or not. This function f is what we call an algorithm or, more formally, a computable function.

More generally speaking, rather than a subset of natural numbers, we can specify any kind of alphabet—that is, any set containing elements distinguishable by some kind of stable regularity that can be abstracted into symbolic form—to provide a basis for the operation of an automaton. The set of production rules for validly combining these symbols into strings plus the symbols themselves (the alphabet) are commonly referred to as the formal grammar of the formal language. This, the formal grammar, specifies all the possible strings and forms that can be produced by or in a given language; hence, this can be mechanically interpreted by an automaton for different tasks, like recognition, classification, or generation of strings [23].

Although different variants of automata may follow more or less complex cases, the point is that every function that can be implemented by an automaton, namely any algorithm, will follow the rules provided by the grammar underpinning its operations (this is a mechanistic process free of any form of decision making).

Furthermore, a function can be said to be computable if it can be effectively calculated by following this kind of step-by-step procedure and yields the correct output for any given input through mechanistic means. In this sense, what Church and Turing (and others) independently demonstrated is that the previously rather intuitive notion of an *algorithm* (or effective calculation or procedure) could be formalized as a sequence of automatized mathematical operations within a formal recursive language [24–27].

Hence, by computation, in this particular sense, we mean the computability of a function. In fact, this is why in [17], irrational numbers such as π and e are considered computable:

$$\pi = 4(1 - \frac{1}{3} + \frac{1}{5} - ...) \qquad e = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + ...$$

Because a number of decimal digits has to be specified for the function, and even if that number is incredibly large, given that the coefficients of the series follow a computable sequence, they can be calculated (see note 2 in Appendix B for a brief note on tractability). Therefore, π and e are not computable as numeric concepts (which would require *understanding* them in a semantic way) but as procedures or programs. In fact, all real algebraic numbers are computable in this sense!

Along these lines, a fundamental property is recursion. Its importance stems from the syntactic operationality that it permits, as it allows self-referential and recursive subsets to be defined from all possible finite combinations of an arbitrary set of symbols (an alphabet) [23,28,29] (see Appendix A for a brief introduction to the notion of recursion in the context of Godel's numbering, and note 3 in Appendix B for some more information).

In this context, it is always important to bear in mind that computation is a mathematical endeavour and that the physical instantiations of computational devices or implementations (natural or artificial) are not really *computational* in nature. While we often use the

term "physical" computation, this should rather be taken as an ascription to some physical system given by their recursive mechanistic dynamics over consistent material properties and relations. These, however, are actually functional descriptions of natural phenomena. Simply put, even if physical *computational* realizations can be described (modeled, analyzed, simulated, etc.) in computational (mathematical) terms, inasmuch as they are materially real, then they *cannot be ontologically computational*.

A good analogy may be Zeno's paradoxes of motion [30], by which Zeno reflects upon the problematic aggregation of an infinite number of finite parts. Let us consider the case of someone walking from point A to another point B. To have traveled the distance from A to B, they must have traveled first the half of that distance ($\frac{1}{2}$), and before that $\frac{1}{4}$, and before that $\frac{1}{8}$, $\frac{1}{16}$, and so on indefinitely. Zeno's point is that even if the individual distances are finite, given that the number of subdivisions is infinite, then it would take an infinite time to actually move from A to B, therefore making motion itself impossible (see note 4 in Appendix B for a brief continuation on this idea).

Likewise, physical computational systems have to be *materially* realized and hence bounded by physical constraints. Unlike a Turing machine working on an infinite tape, their memory is restricted and their time limited. It follows that any system that is Turing complete (like most modern programming languages or Conway's Game of Life, for example), would be capable of computing what a Turing machine can compute if it were given infinite resources but no more than that [27].

This difference in applicability between mathematics and the natural sciences is probably the source of many contemporary confusions and unknowns. In fact, unlike mathematical undecidability, physical undecidability has not been proved to exist, and the main reason for this is that every case relies on some form of infinity [27]. Undecidability proofs in physics are mathematically rigorous, but whether they apply to physical reality depends on whether the idealizations they use (infinities, continua, or unbounded precision) are physically real. In simple words, what we really have proven is mathematical undecidability in physical models and not undecidability in nature. This has led many researchers to think that every process within our physical universe can be abstracted as a computable function [27,31] (we will return to this later).

Building on the thought experiment proposed by Van Gelder [32], if we were to consider a Watt governor, would we say that it is performing computations when regulating pressure and heat or just following the simple dynamics dictated by its physical design? Most would agree it is the latter. On the one hand, conflating a system's behavioral and computational descriptions seems to arise from ascribing goals to processes that presumably do not have any. On the other hand, and more importantly, when we speak of something being computable, we do not mean that its physical implementation is literally performing computations—perhaps apart from some actual modern computational device. We intend that its operation can be modeled and therefore abstracted mathematically as a sequence of concatenated operations that do not require a mind (i.e., someone thinking or making decisions), even if the actual implementation of such mathematical descriptions may be far more inefficient! (As a matter of fact, it has been shown that a hypothetical implementation of an algorithmic Watt governor would perform much worse than just leaving the device to operate according to its physical design alone [32–34].)

To be clear, whereas a system being modelable is an epistemological claim, concerning the system being formalizable in computational terms without the system being a computer, the alternative idea (often expressed by the term "physical computation") usually implies an ontological claim, namely that such systems *implement* computational processes and therefore that the system's causal structure *realizes* computable functions. In this paper, we intend to advocate for the former.

In the next section, we will review an influential hypothetical implication stemming from the concept of computation, namely the Church–Turing thesis.

2.2. The Church–Turing Thesis

In simple words, the idea that any general form of computation will be equivalent to any other, and therefore to a Turing machine, is what is known as the Church–Turing thesis. Although it is known as a *thesis*, it could be said to be rather a *conjecture* about the nature of computation, as it is unprovable in a mathematical sense [34,35]. It is unprovable because the idea of any general form of computation is essentially the same as anything that can be effectively calculated, which is still an intuitive notion; there is no formalization that encompasses every possible kind of computation. Nevertheless, while unprovable, it is at least heavily supported by the fact that every general model of computation known to us has the same computational power [23–25,27,35].

The development of these ideas went as follows: A formal definition for the notion of algorithm was introduced by Church [20] and Turing [17] through lambda calculus and Turing machines, respectively [26], entailing that they are equivalent. Subsequently, Church [21] and Kleene [36] proved the equivalence between λ-calculus and general recursive functions, proving the so-called Church thesis, the claim that the class of functions which are computable by a finite mechanical procedure is the same as the class of general recursive functions [37]. Moreover, after this several independent models of computation (such as Post systems, register machines, type-zero or unrestricted grammars, and cellular automata) have all turned out to be equivalent [23,25,35,38–40], leading to the conjecture that any other general form of computation will also be equivalent in expressive (computational) power, also expressed as: $\mathcal{L}(TM) \subseteq \mathcal{L}(UG) \subseteq \mathcal{L}(\mu RF) \subseteq \mathcal{L}(TM)$.

This is the same as saying that Turing machines can be imitated by grammars, grammars by μ recursive functions, and μ Recursive functions by Turing machines. Hence, that they can compute the same functions and recognize the same formal languages [41]. In this sense, assuming the premise is (presumably) true, the central question is whether what the human mind does is really—or can be understood as—a form of computation or not, and *whether every mental feature is computable or not*. This is the fundamental issue.

Often, we get the mistaken impression that computers can compute anything, which certainly is not the case (not even with infinite time and memory). This is in fact a central insight from the works of Godel and Turing; at the center of the Church–Turing thesis, not only can every form of computation *do* the same, but they will all be limited to the same power [24,27,29,35]. Indeed, as we have mentioned, the functions that Turing machines can compute are a restricted set of all functions which only becomes more restricted in the case of physical instantiations [42] (see note 5 in Appendix B for some other views).

Another central aspect of the Church–Turing thesis is the functional nature of the mathematical descriptions. A proof for this is that systems that can be physical instantiations of computation are multiply realizable; consider, for example, Babbage's analytical engine, in contrast to modern electronic devices [18,19]. The same applies to chemical or biochemical cases [7,43]. In this sense, given that computation pertains to abstract descriptions or models of phenomena in a physical realm, a point of debate is the possible role of this physical substrate as a hypothetical *unaccounted* component of the mind, therefore, as something that cannot be captured by the kind of abstractions that computation permit, even in spite of its extraordinary generality.

It is important not to conflate this with the difference in the specific instantiations. In fact, along the same lines as the Watt governor example described earlier, we could say that even for a universal Turing machine, it would be impossible to perform the atomic steps needed in a parallel system, in which all the cells are updated simultaneously (in

contrast to the serial Turing machine), or that of an enzymatic system, where the atomic steps involve operations such as selective enzyme binding. This, however, does not mean that a universal Turing machine is incapable of calculating their behavior or simulating both parallel and enzymatic systems, at least to a certain level of accuracy. The algorithmic version of the Church–Turing thesis states that this is true of every algorithmic system, even if highly likely in quite convoluted ways! [11,19,35]. Therefore, *the problem is not whether the brain or body is a computer or not, but whether the mind that we believe emerges from it can be fully characterized through computable functions*.

The relevant point is that insofar as the notion of computability can be applied to any material entity and process, the existence of non-computational phenomena is not something we can take for granted. In this sense, a classical example of non-computable behavior is randomness. However, given that randomness is fundamentally non mechanistic, it does not seem to be amenable to a formal proof in a logical or mathematical sense [44,45]. Thus, the non computability of some dimension of our material universe may also be seen as a strong conjecture [19,46], very much akin, albeit opposite, to the Church–Turing thesis.

In this context, the idea that certain physical phenomena might be instances of non-computable functions and therefore outside this restricted set has been proposed as a fundamental feature of life [47–52] and mind [12,14,53,54]. While this opens a speculative dimension that cuts both ways, we will work under the assumption that some phenomena, like radioactive decay, are evidence of physical non-computability related to unpredictability [46] (we will return to this topic in Section 4).

Note that unpredictability is considered non-computable when no finite algorithm can, even in principle, reliably determine a system's future states, typically because the system's behavior encodes undecidable problems, and depends on non-computable real numbers requiring infinite precision or produces algorithmically random outputs with no shorter description than the sequence itself. In such cases, prediction would require solving problems known to be uncomputable or accessing information that no algorithm can finitely represent. This is different from ordinary epistemic unpredictability, where we simply lack enough information.

Before moving onto the next section, it is important to stress that the fundamental dilemma regarding the Church–Turing thesis is twofold: (1) whether Turing machines are computationally equivalent to the human mind, and (2) more importantly in our opinion, whether every form of cognition (and therefore the human mind) is actually something that can be fully described through computational abstractions.

To summarize: Are the mechanisms from which we assume life, cognition, and minds emerged all a general form of computation? If so, and hence of the same underlying nature and with the same limitations as any other physical realization of general computation, then human traits such as phenomenological experience and agency should be perfectly possible in artificial systems, theoretically at least. If not, then how could we make mathematical and scientifical sense of this kind of physical phenomena in cognitive organisms?

In the following section, we will further explore this issue. We will begin with a brief historical review to portray how the association between life and cognition has influenced our understanding of what the mind is and then move into some of the most influential current positions in cognitive science, within the context of the overarching topic of cognition and computation.

3. Cognition, Life, and Computation

In this section, we present a critical discussion of the ideas in cognitive science, from cognitivism, to enactivism, to current positions like the free energy principle and radical enactivism. We illustrate how the notions of cognition (and therefore of the mind)

are unavoidably linked to computational conceptions and why it is so hard to make sense of hypothetical alternatives, which may appear more natural for notions like life and mind.

Basically, cognition is the set of processes by which a system acquires, transforms, stores, retrieves, and uses information to guide action, adapt to its environment, and maintain its own organization, including perception, memory, decision making, and learning [34]. Depending on the theoretical framework, it may be understood as something that occurs not only in human minds but also in animals, minimal biological systems, or even artificial systems so that whenever there is a consistent set of mechanisms allowing structured information processing, it supports adaptive behavior [2,55–57].

We could portray the current overall view of cognition—nuances apart—as a form of organic intelligence or organic computation. However, there are some core theoretical assumptions that are still deeply controversial and which underlie different interpretations, especially regarding the idea of the mind, insofar as a cognitive function, or features [56,58–60]. This historical survey aims to underscore the depth and relevance of these differences, explaining how they led to the current state of affairs.

3.1. Mind as Intelligence: Computers and Organisms

On the one hand, cognition has a strong association with the notion of mechanism and therefore with the broader background of computation. In fact, artificial intelligence has been one of the central fields of cognitive science since its origins, originally entailing a view centered around ideas like models and representations, namely cognitivism. On the other hand, there is a natural relation between cognition and life itself, engendered by the fact that we can trace intelligence to extremely primal forms of life. Thus, a contrasting view emerged in which cognition and life are fundamentally intertwined: enactivism.

As we will see in this section, these strict initial differences have mostly faded away or resulted in mixed approaches, giving rise to a variegated landscape.

3.1.1. Cognitivism

Cognitivism is a label for the foundational era of cognitive science that started with the cognitive revolution (mostly in opposition to behaviorism), a computational approach to intelligence and cognition underpinned by the idea of syntactic manipulation of symbolic representational structures (mental representations) in an algorithmic manner [34,61].

Roughly speaking, the relation between mental representations and the computational approach in cognitivism can be traced back to three main sources: (1) psychological studies, which introduced the notion of mental representation as a natural solution to the limitations of mental capacities [62,63]; (2) developments in artificial intelligence, stemming from the Church–Turing thesis, along with work in cybernetics on information and representations in the brain [22,64–66]; and (3) the development of formal syntax of languages by Chomsky [38,67], which settled the basis for automata classification in terms of grammar production and recognition and of language as something algorithmic (i.e., mechanistic).

All of this translated into the idea that the brain was a particular physical implementation of computable functions, in which mental representations played the role of symbols and where logical operations were physically carried out by neural machinery [68–71]. Overall, this implied that consistent perceptual faculties must be underpinned by some physical mechanism that overcomes the physical limits of the system (e.g., memory to remember past events or models of the world to navigate a city).

The emblematic development within the cognitivist framework was the physical symbol system hypothesis (PSSH) [70,72]. The core idea of the PSSH is that there are universal limitations to the kind of system that can display intelligence. First, given the

physical nature of our universe, even if multiply realizable, a system must be physical. Therefore, whatever its properties may be, they will be enabled and constrained by natural laws. Secondly, given that there must be a consistent correlation between physical reality and the physical states of the system guiding its behavior, any such system will need to instantiate physical symbols to consistently represent and safeguard these external-internal correlations, allowing perception and generative production.

Hence, material intelligent systems, including human brains, would be physical instantiations of another general form of computation as put forward by the Church–Turing thesis. Furthermore, if we assume that any general form of computation will be equivalent to any other, then any realization of the same principles will be functionally equivalent and hence underpinned by the same fundamental properties stemming from syntactic recursion and symbolic representation. And given that this extends to every form of computation, then any physical instantiations of the same principles will display the same capacities and will be bound to the same limits. This being the case, the difference between mental machines and other computers would be one of machine complexity and not of kind.

3.1.2. Autopoiesis

Briefly put, autopoiesis refers to the specific intrinsic property of living beings by which a network of interrelated elements recursively produces the same elements that constitute the network that produces them [73] (p. 67) (see Appendix C for a definition).

Also along the lines of the cybernetic tradition [65,74–79]—that is, of control independent of the substrate—Maturana and Varela [73] abstracted functional principles and defined autopoietic systems as machines that, given their particular organization, are able to continuously self-produce in a physical space by transforming external elements into their own components while subordinating all structural changes to the conservation of the same organization, giving rise to an *autonomous identity*, namely an observer-independent organized unity capable of intrinsic behavioral determination subordinated to its own subsistence and self-production [73] (p. 69). In the case of the living cell, for instance, there is a network of reactions that produces the membrane which, in turn, encloses the components and their medium, facilitating the production of the membrane itself [73,80].

This circularity is, in fact, an essential notion within the broader framework of theoretical biology [11,49–51], particularly for the (M,R) systems formally introduced by Rosen [47,48] as well, to the point that it has been proposed that autopoietic systems are a subset of (M,R) systems, with both being non-simulable by Turing machines [50,81].

As noted by Virgo et al. [82], whereas cognitivism can be seen as a product of cybernetics research in the sense of regulation and control through a model of the world [78], autopoiesis can also be seen in this way but in the sense of self-organization [83] and regulatory feedback processes [75]. Indeed, autopoietic organization may be portrayed as a persistent, self-organized set of interwoven feedback loops somehow similar to Ashby's ultrastable system [66] with only one variable to sustain: its organization [73,84].

Together with the main biological thesis, the theory presented implications relevant to cognitive science, such as the concept of cognitive domain, that is, the set of all structural modulations that a system can perform without disintegrating. From this point of view, biological cognition is equivalent to behavior, inasmuch as system responses are a manifestation of a particular form of *knowledge* or intelligence about the adequate modulation of external circumstances, based on what the system can perceptually distinguish [73,85–87]. Cognition, then, would be the capacity of the system to adapt to or *structurally couple* with environmental contingencies without losing its organization [85,88], something achieved even by a single cell without the need for complex models or representations.

Autopoiesis portrays a view of cognition as a biological and structural *blind* intelligence that differs greatly from cognitivist approaches, for which symbolic representations are the basic substrate for any form of intelligence. In this sense, the theory is relevant because it posits a purely mechanistic view of cognition rooted in biological dynamics [85,89] nonetheless imbued with a subjective perspective [90] and a non-symbolic form of intelligence, hence posing an anti-computationalist view.

This opened up an alternative path for interpreting the relation between cognition and computation, one radically different from intelligence conceived as a solely human feature and where deterministic coherences would give rise to autonomous intelligent behavior, life, and cognition. In the process, it cast doubt on the idea of brains (insofar as physical computers) as the only source of intelligence worth considering.

3.1.3. Autonomy and Enactivism

Overall, autonomy is a more general and abstract conception of the principles behind autopoiesis, supported by the introduction of the concept of *organizational closure* to designate cases of recursively defined organizations, represented as closed sets in which operations within the set can only result in elements belonging to the same set [91–93].

In this more abstract connotation, the domain is said to be closed because during its operation, from a whole space of possible states, the system will only occupy a limited subset in which its viability conditions are not transgressed, therefore transitioning from viable states into (new or not) viable states and defining an operational subdomain [91,92]. This recursive materialization—operational closure—would then be the fundamental dynamic property necessary for any form of autonomy, and autopoiesis is a particular case of autonomy (i.e., autonomy of the living cell; see the glossary in Appendix C for disambiguation).

Contrary to cognitivism, in which the body and mind are seen as clearly distinct, analogous to hardware and software, autonomy presents a less categorical view. Whereas computational artifacts interact with the environment by being fully specified by the input/output relations predefined by some external design, for autonomous systems, what may count as an *input* will depend on the organization and history of the system [73,94], thereby dissolving the strict separation between machine (physical object), programs (plans for computation), and data (numerical or symbolic input).

Later, in *The Embodied Mind* [95], autonomy was further developed within a proper cognitive science framework and presented alongside other concepts, like embodiment and enaction, providing a plausible explanation for how cognitive systems (e.g., the nervous system) materially encode their organizational specifications (*transition functions*) and their history of interactions in the world (memory) by reshaping their structural configurations [55,96]. Accordingly, the specific sensorimotor capacities given by the system's material realization (its body) and its particular environmental conditions would co-determine its particular cognitive interpretation, known as an *enaction* [95].

Roughly speaking, the source of the differences between the cognitivist and enactivist approaches to cognition lies in the concept of embodiment. In enactivism, the body is conceived as the concrete domain of the cognitive phenomenon—rather than a shell for the mind—in which adaptive environmental interactions are *realized* without the need for models or representations [1,55,95,96].

Whereas autopoiesis is anti-computationalist, the more mathematical presentation of the notion of autonomy favors computational description, at least insofar as mechanistic series of structural transformations. In fact, a proof of concept for autonomy is presented in [94,95] by means of a toy model named Bittorio, a ring-like, organizationally closed cellular automaton that transitions (adapts) between stable dynamical states when exposed to specific external conditions.

While it simply alternates between a pair of configurations (through a rule that enables it to recognize even or odd external sequences), Bittorio can be understood as displaying a minimal form of system–environment coherence, whereby the interaction of the system and environmental states produces a syntactic co-determination of its behavior (structural changes) (in other words, a minimally (enactive) autonomous system).

Ideas along these lines, particularly the notion of autonomy, were well received and developed by more concrete research on bio-inspired computation and robotics [97–104] and later by artificial life [58,105,106]. Insofar as these approaches combined a non-representational perspective with a strong drive for computational implementations (i.e., as physical instantiations), they may be seen as the culmination of the cybernetic search for intelligence in terms of natural mechanisms [107,108] and the mark of an applied synthesis of the two theoretical approaches stemming from computationalism and biology of cognition.

The achievement of artificial minimal cognitive behavior was concrete proof that formal characterization of biological-like features could explain intelligence in terms of system's dynamics; *cognition is not computation, as originally purported, but it is computable, insofar as embodied organic machinery*.

This triumph, however, brought new questions into play. Will AI systems be capable of thinking like we do? Will they acquire, at some point, valenced emotions that guide their intelligent behavior? Will they have a mind?

To some extent—even if only principle—the original question about the relation between cognition and intelligence had been answered, leaving, however, a sense of *incompleteness* behind.

As a matter of fact, we can say that these advances (theoretical and applied) are what triggered the shift from classical to modern AI and from intelligence to agency and phenomenological experience as the distinctive features of the mind (see Figure 1 for a schematic presentation of these ideas).

3.2. The Mind Beyond Intelligence

Is this new conception of cognitive systems (enactivism) so different from the classical physical symbol system hypothesis? Or could it be seen, rather, as a form of organic computation (describable as recursive mechanisms operating on physical symbols or distinctions), which follows mechanistic transformations obeying material selectivity, and a particular kind of system–environment interaction and memory handling?

Let us consider Bittorio [94] as an example. The system provides a formal characterization of the idea advanced by Maturana [109], Maturana and Varela [73], and Varela [91], namely the origin of subjective, albeit through a purely syntactic interpretation. More specifically, given that the system has a degree of selectivity (i.e., it responds to some perturbations but not others), it can be said that it *interprets* its environmental background, in the sense of specifying structural changes with respect to some regularities by covarying with them [94,95]. This is another way of saying that what is being structurally embodied are these structural correlations and hence some form of *information* or knowledge about them. In other words, the fact that this information is subjective does not seem to preclude the dynamics of the system from being described in computational terms.

In fact, a frequent confusion derived from the original enactive ideas has to do with the different types or levels of autonomy. Throughout most of *The Embodied Mind* [95], the main type of autonomy being discussed and argued as capable of supporting experience, as well as meaningful forms of cognition, is the autonomy of the nervous system. But the notion of autonomy itself is presented as something much more general [96]. For instance, the iconic case study supporting the enactive approach to perception is related to neural, linguistic,

and cultural correlates of color perception [95] (pp. 157–171) within a specific human context. Bittorio, on the other hand, a minimal, closed cellular automata governed by a simple transitional rule, was equally posed as an example of an autonomous system [95].

Figure 1. A broad conceptual map contextualizing the main research positions presented in Section 3.1 within the research framework underpinned by the idea of mind as intelligence. Above in orange is the inceptive and fundamental work in cybernetics, looking for intelligent mechanisms for control of behavior based on both biological and computational approaches before a division of focus occurred. In blue (left) are approaches from a computational framework, based on the the underpinning constructs of representations forming models of the world. In green (right) are anti-computationalist approaches based on theoretical biology. In orange (bottom) are synthetic approaches embodied by bio-inspired artificial intelligence and robotics, with physical (computable) implementations working without predefined models or representations. Some seminal works are included within each box for reference (some titles were shortened for reasons of space). See the main text for further details.

Whereas Bittorio is presented as an autonomous system capable of enacting its environment by structurally covarying with it, it is made clear by [95] that a minimal model like this is obviously devoid of *meaningful* forms of experience. Thus, any non-organic

autonomous system would be capable of identifying and enacting environmental features, *bringing forth a domain of significance*. However, in spite of the richness of such terms, this only means having consistent structural selectivity to external regularities and a closed organization that allows an associated coherent modulation given by intrinsic mapping-like dynamics, things that are amenable to computational characterization. In other words, Bittorio is a syntactic processor that is missing something that minded beings have—or are supposed to have—even if we cannot really say what this is.

This gap (between Bittorio and human experience) is a metaphor for the source of the conflict in enactive cognitive science, which has given rise to new approaches to cognition, or computation and the mind.

Whether it is a matter of degree (such that more intelligent systems can eventually develop agency or phenomenological experience) or kind (in which case we are missing some fundamental property responsible for this leap to the mental) is still one of the most debated issues in current cognitive science and other disciplines, such as philosophy of mind, neuroscience, and artificial life [4,60,110–113].

As we shall see now, it is these sorts of questions that have fragmented the landscape into different positions [96,114,115] while also leading to a return to computationalist views on cognition, this time motivated by biological foundations (see Figure 2 for a rough summary of this section.

3.2.1. Bioenactivism

Enactivism can be considered to be an attempt to reconnect formal developments in cognitive science with the elusive problem of what the mind is, in terms of phenomenological experience. This is particularly true for bioenactivism. Variously called "autopoietic", "traditional", or even "hardcore" enactivism, bioenactivism bears a fundamental commitment to the idea that life or, maybe better, the striving for survival is the ultimate source of the mind [116–121]. Central to this strand is both the philosophical influence of phenomenological authors like Husserl, Merleau-Ponty, and especially Jonas [122,123], whose ideas set the basis for the life–mind continuity thesis (i.e., the idea that mind emerges from the living) and also the introduction of new concepts and reformulations of problematic assumptions from the original ideas of autopoiesis [96,124,125] while still drawing heavily on the ideas in [95].

Subscribers to stronger or more traditional versions of the life–mind continuity thesis hold that research in autonomy forms a basis for understanding the emergence of a mental *experiencer* from subjectless systems [96,119,126–128]. Put another way, the constant threat of disintegration would bootstrap a need for and a course of action into self-individuating systems, either toward energy sources or away from potential hazards. Therefore, insofar as this displayed intelligence involves the (continuation of the) existence of the system itself, living organisms will be a specific class of autonomous systems, endowed with the built-in imperative to keep existing and hence with teleological behavior [117,129,130].

A fundamental guiding assumption for bioenactivism is that autonomous sensorimotor systems are *agents* and thus capable of actively and asymmetrically determining their behavior [60,131,132] and, in this respect, being particularly critical of purely sensorimotor-based accounts of experience [1] (pp. 29–32) and the notion of a blind and brittle structural coupling in autopoiesis [124].

Adaptivity [124,133] (see Appendix C for a more a more formal definition), along with the notion of precariousness [1,134], in this context, becomes key to the attempt to fill the spaces in the work of Varela et al. [95] by supplementing the notion of autonomy with intrinsically motivated *adaptive* behavior. More specifically, given that the viability conditions for autonomous agents are rooted in the nature of their specific operational

closure, the intrinsic source for regulation of their behavior would be to remain within the viability bounds [1] (pp. 120–121). This would be accomplished by means of their adaptive capacity by actively monitoring, evaluating, and regulating their own states and their coupling with the environment with respect to their intrinsic normativity [1] (pp. 122–123). In other words, cognition is, or is realized through, a *sense-making* operation that requires autonomy plus a graded normative adaptivity underpinned by viability conditions, and this brings forth a phenomenological world of affective forces [120].

At this point, the problems resurface, because even if the overall framework offers a robust bottom-up elaboration, its commitment to the premise of teleological intentionality entailed by the autonomy of living beings inevitably leads to arguments being cast in such terms [125]. In fact, once we reevaluate computational descriptions, we can see that where such modulations occur, we can expect the sorts of reliable regularities suitable for survival, describable as mechanisms, and mappable onto input–output relations.

For example, let us consider the idea of adaptivity as a precondition for sense making. If evaluations are made by computable processes (by means of coherent structural changes, acting as recursive mechanisms), then the problem is that structural changes are supposedly evaluated in semantic or meaningful grounds, but then, any sort of evaluation would require more structural changes itself, thus implying new evaluations, and so on ad infinitum.

In simple words, if monitoring and evaluating take place by following purely syntactic, computable processes, then could we still think of a mind? If, on the other hand, we presuppose some semantic or phenomenological component, then where would this come from: outside or beyond cognitive or computable mechanisms?

3.2.2. Radical Enactivism

According to the radical enactive cognition (REC) approach, notions like evaluation, sense making, or generation of meaning in basic minds have unavoidable representational requirements which are not properly justified [135–137]. REC posits a vision not only devoid of representations for simple minds but of meaningful intentionality as well (as living Bittorios, so to say). Accordingly, cognitive behavior is characterized as a strict organism-environment informational covariance that is solely the consequence of biological evolutionary forces, along with the adaptive ontogeny characteristic of living forms. Hence, there is no need for an agent "inside looking out" guiding behavior by making judgments, or as Hutto and Myin put it, *"Why should simply having a biohistory [...] be thought to entail the existence of any kind of semantic properties?"* [137] (p. 111).

To fill the void left by removing the sense maker from the equation, REC has introduced the concept of *ur-intentionality*, a hypothetical, information-based, primitive, and goal-oriented tendency to action. In simple words, ur-intentionality can be depicted as a non-contentful correlation of the internal states of an organism and its environment, quite close to the idea of blind organic machinery [115]. Thus, similar to how the rings in the trunks of trees are the manifestation of a system–environment coherence (a biohistory of correlations), simple organisms would correlate behaviors with particular external circumstances without the need for a mind or an observer.

Roughly speaking, the argument is as follows. Let us consider the idea that there is no global mind emerging from life itself such that at least minimal living systems can be conceived of as incredibly fine-tuned evolutionary ensembles of individuals (a sort of colony) that have evolved to act together as a whole but where there is no collective overall sentient system. Why do they remain together, or even better, why or how did they undergo such a process of amalgamation? It is simply because it was beneficial to all the previously independent individuals, to the extent that they became not only strongly

interrelated but also heavily interdependent. Therefore, while a flock of birds can split and join again through successive instances of self-organization, and they can still live as individual systems, the former type, after some critical no-return stage of evolutionary amalgamation, cannot.

An organismic convergence of this kind must not only bind sub-organisms together physically but, more importantly, behaviorally. In this sense, the primal natural selectivity of the individuals must give way to a combined selectivity that progressively overrides or coerces them (i.e., the emergent organization), establishing an organismic intrinsic coherence that results in the appropriate behavioral patterns that we are able to observe from the amalgam. It is this emerging selectivity which becomes specific to the survival of the whole over that of the individuals which would bootstrap ur-intentionality, because the behavior of the system is now objectively directed toward the distinctions arising from the emerging selectivity. There would not be a need for a global mind whatsoever, and there would not be any real intentionality either not because there is no directedness toward or about an object, but because there is no observer that could hold an impression about such an object.

By stripping autonomy of the idea of some property "pulling the strings" behind sense making, REC offers a principled and naturalized understanding of the origin of cognitive behavior (as of sense making). Unfortunately, it does not provide us with a concrete explication of the mind and even less of phenomenological experience in an analogous manner. And it is at this point where the problem becomes profoundly evident. By voiding autonomy of life from a unified mind, which could become the observer, radical enactivism leaves us almost in the same place as before [138,139].

In this sense, there is no clear bridge from a hollow organismic machine as depicted by REC to some evolutionarily subsequent cognitive property that could turn ur-intentionality into meaningful intentionality, because even if we were to consider a perfectly or fully amalgamated system (hence with absolute global selectivity and directedness, as in autopoiesis), then we would be taken back to the original conundrum. Put differently, a better understanding of the mechanisms underpinning life not only does not seem to be helpful to understand the relation between life and mind, but it appears to increase the gap between them.

3.2.3. Back to Autopoiesis

From an even more *radical* stance, and aligned with the strict naturalism defended in the original ideas from autopoiesis [73,85,88,140], the autopoietic theory of cognition (ATC) posits a view of cognition as structural processes that do not necessarily involve phenomenological properties (i.e., meaning, sentience, consciousness, etc.) until the much later appearance of higher-order social capacities, specifically language [8,57,59].

The problem would be teleology (i.e., purposeful or goal-oriented action) and underlying notions like adaptivity, normativity, or agency, thus implying a kind of systemic control and regulation based on properties beyond natural phenomena (thus beyond their mechanistic biological nature) [59,125] and, therefore, beyond scientific descriptions.

Furthermore, it is suggested that directedness, the property attributed to a goal-oriented organism as a whole, is rather an illusion created by our anthropocentric way of conceiving the world, like previously thinking the Sun moved around the Earth or that of a virus 'looking for hosts to invade'. In this way of thinking, life would not be any different; it may seem to us that living systems act toward their environments but only because it is natural (for us) to perceive them so, given our preconception of life, aside from the intractable complexity of their internal dynamics [8,125,139]. In other words, the construction of any form of coherence between the recursive network of processes and

whatever is not the network itself amounts to a blind convergence by blind mechanistic operations, hence along the same lines as Maturana and Varela [73].

Basically, ATC sees in enactivism a form of wishful thinking about simple organisms, *as if* they were something more than highly complex evolutionary machinery. As a matter of fact, by following this line of reasoning, the ideas posed by REC can actually be taken further not only to characterize behavior as devoid of intentional semantic content but also of any *unnecessary* notion of global directedness and hence any kind of intentionality as well [141]. Simply put, there is no mind we can refer to, no active or passive experiencer, and no ghost in the machine at all.

A different 'return' to autopoiesis has been put forward by the basal cognition approach [2,3,89,142]. Inspired by research in fields like comparative psychology and cognition as well as bio-inspired robotics [98,143–145], basal cognition proposes that cognition starts with life, and it can be found in any kind of organism (e.g., bacteria, plants, or slime molds), thus having no strict need for a nervous system [146].

Whereas this view is akin to enactivist and autopoietic takes on cognition, the framework follows a biological empirical account and does not commit to a strong philosophical stance about the nature of cognition. Simply put, there is no a priori rejection of computational ideas (see, for example, [147]). As a matter of fact, research in basal cognition is deeply linked to research on morphological computation [148,149], given that both aim to understand the fundamental features of minimal (basal) cognitive traits. In the latter case, this is often with actual implementational purposes [150–153].

Following the principles laid out by bio-inspired robotics, morphological computation has analytically unraveled the enactive idea of embodiment in terms of intelligent behavior. Indeed, this is to the point where instructions can be communicated as or programmed into ensembles of genetically created cells, therefore making multiple realizability a strong theoretical option once again [149,152,153].

Like most biologically motivated endeavors, these approaches still attribute something *special* to life, although in this case, maintaining a distance from strong phenomenological associations. Specifically, following Dennet [154], there is purported agency in terms of goals and goal directedness [149,155]. In this sense, in spite of sharing ideas with enactive positions, both basal cognition and morphological computation presuppose quite distinct mechanisms underlying cognition, amenable to computational ascriptions, and therefore portraying a different kind of *functional* mind.

3.2.4. Active Inference

Perhaps ironically enough, this return to autopoiesis can be seen, within a broader context, as a return to a new form of computationalism. Indeed, in most contemporary cognitive science, this seems to be the case, especially given the importance of neuroscience and artificial intelligence nowadays. Active inference is part of a layered framework that also includes the Bayesian brain hypothesis, predictive processing, and the free energy principle, all of which share underlying principles.

Basically, the Bayesian brain hypothesis and predictive processing are theories about the brain; the former proposes that brains instantiate probabilistic beliefs that are updated following Bayes' rule, and the latter investigates how this might actually be implemented, assuming that the brain works as a hierarchical predictive machine that minimizes prediction errors by contrasting its predictions against actual sensory signals and then updating its *model* of the world by adjusting the connectivity among neural layers [156–158]. Although interesting, these approaches may be too narrow considering the scope of this work.

Conversely, the free energy principle (FEP) is the generalization of these ideas, which proposes that every living system minimizes surprise (variational free energy) in order to

maintain its integrity, a cognitive operation that can be traced to the minimal biological case, i.e., autopoiesis [56,159]. Active inference, in turn, specifies how these principles may be implemented through perception and action processes by diverse organisms [160,161].

To a great extent, active inference can be seen as a synthesis of cognitivism and enactivism. On the one hand, its framework is fundamentally built on top of ideas like models, information, and representations. On the other hand, by reinterpreting computational ideas in light of an enactive understanding, it has produced some incredibly important insights into problems posited by enactive accounts [160].

We could even say that it presents a new form of enactivism, one that aims to overcome the *spookiness* arising from the strong "life = mind" thesis, supported by rigorous mathematical and computational foundations. For example, the enactive view of active perception was only fully exploited (insofar as representations within an extended cognitive window) by the framework of the FEP and active inference [56,157,162,163].

Another key element is the incorporation of a top-down component, which was mostly alien to enactive developments in cognitive science [11]. In fact, this top-down component would endow the system with *predictive* faculties, which would be the cognitive source of the *temporal depth* that characterizes our conscious experience [110,112].

Again, despite the importance given to temporality and phenomenology by enactivist accounts [164–167], such ideas were rarely formally developed. Active inference, however, departing from the initial neurophenomenological project [167,168], accommodated Husserlian phenomenology in a seemingly better way than enactivism itself [112,169–172].

Part of these advances crystallized into computational phenomenology, an approach that uses computational modeling to formally explore the structure of subjective experience [173,174]. Instead of treating phenomenological experience as introspective only, computational phenomenology considers phenomenological features (temporality, intentionality, or agency) to be modelable features of information-processing systems.

Unlike traditional phenomenology, which is purely descriptive and philosophical, approaches of this kind, part of the broader framework of naturalized phenomenology, aim to reconcile phenomenology and empirical science by mapping behavioral, cognitive, or neural processes to the mathematical structures representing the dynamics of experience [172,173,175]. Do note that this is different from research on the neural correlates of consciousness [176], which is specific to neural activity and does not provide, nor aim to provide by itself, an explanation in deep phenomenological terms.

In this sense, however, while the goal is not to reduce or instantiate phenomenological experience into computational machines, it is unclear how to avoid a reductionist take [177]. Inasmuch as the models serve as hypotheses, there is an inescapable assumption about the computability of cognitive phenomena. This is not tied to the premises of active inference nor to computational phenomenology, but it is a fundamental limitation of our research methods.

In simple words, if cognitive hypotheses cannot be presented computationally, then they cannot be corroborated or falsified, and if this is the case, then they are not within a scientific domain any more. Put against this, what kind of formal hypotheses could we actually investigate otherwise?

In this section, we have tried to give the reader a broad overview of pertinent positions in cognitive science. While there are no universally accepted ways of categorizing the different approaches—indeed, it can be argued that attempting this is problematic [1,55,57]—the "position" criteria used in Table 1 is one possible way of summarizing some of the key properties of the approaches included in our review which readers might find helpful.

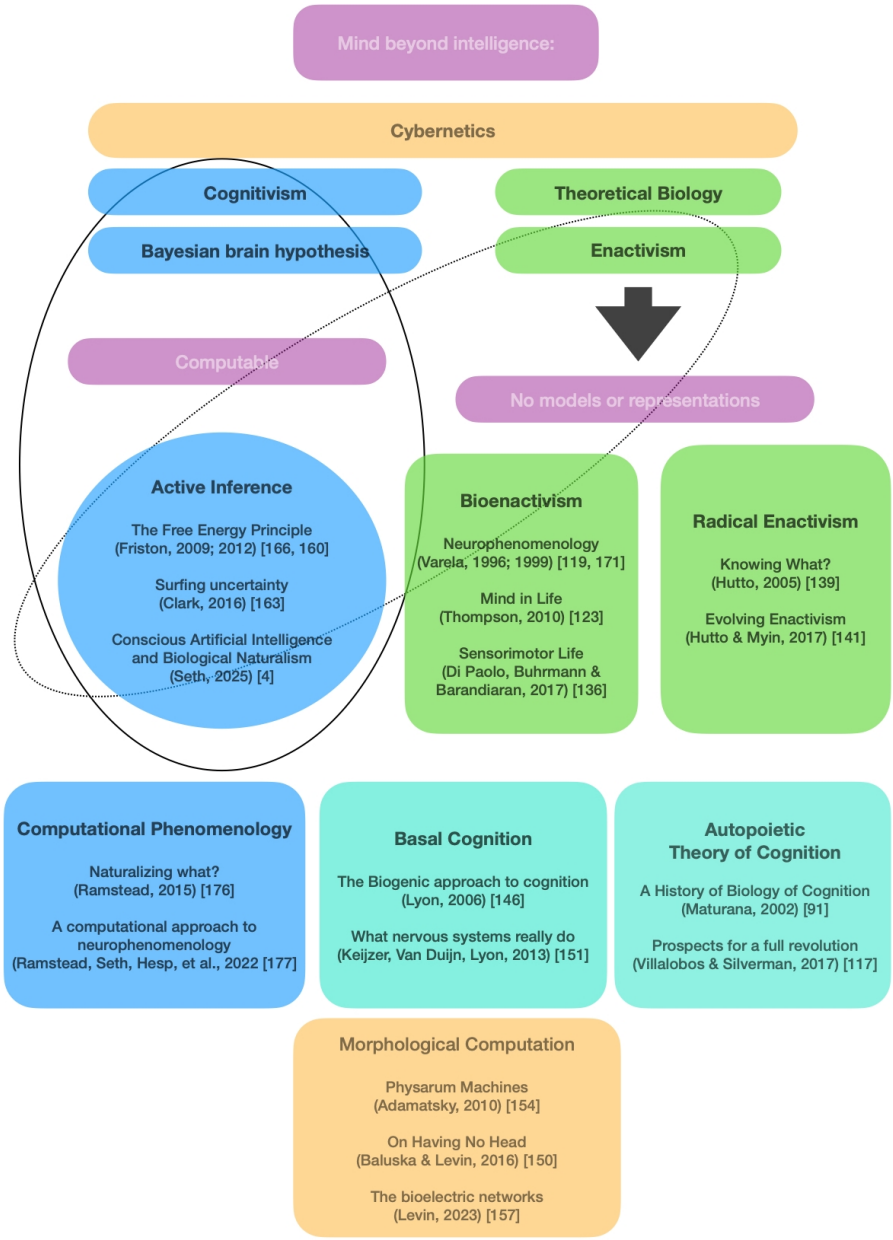

Figure 2. A broad conceptual map contextualizing the main research positions presented in Section 3.2, based on the idea of the mind as something *beyond* intelligence. Cybernetics (orange, top), cognitivism (top left, blue), and theoretical biology (top right, green) are the main theoretical sources for current views (for a more detailed diagram about this, see Figure 1). Bioenactivism and radical enactivism (green, middle right) have followed the anti-computational tradition of enactivism but have not been able to provide solid empirical grounding for agency or phenomenology. As a synthetic response, active inference (middle left, circular blue) has integrated ideas from cognitivism and enactivism. Interestingly, "descendant" accounts of enactivism (calypso, right) have "returned" to—or reinterpreted—autopoiesis while distancing themselves from phenomenological ideas. Conversely, active inference has actively engaged with this research, with computational phenomenology (blue, bottom left) being the best example. Somewhat analogous to bio-inspired robotics, morphological computation (bottom, orange) plays a synthetic role, focusing on a form of agency, or goal-directedness, that does not invoke phenomenological decision making. In general (except for ATC), it can be said that current advances agree on the idea of life as something different in kind but disagree on the property (e.g., sentience, agency) conferring life such a status. Some seminal works are included within each box for reference (names may be shortened for reasons of space). See the main text for further details.

Table 1. Table displaying essential commitments of the different positions reviewed in Section 3. In the first row, while bioenactivism is fundamentally non-representational, newer positions place harder requirements for a unified observer (i.e., a mind), so much so that mental representations are an important component of active inference. The same pattern can be observed in the following lines (normativity, teleology, and autonomy), which showcase a deeper change in cognitive science toward more *rigorous* or scientific and positivistic grounds, in this specific regard. Finally, whereas bioenactivism is rather strictly anti-computationalist, REC ascribes to the use of information as a valuable tool, even if it rejects cognition as contentful symbol manipulation. Along these lines, ATC's ideas are mostly amenable to those of physical computation, while active inference basically describes cognition as sets of computable functions. See main text for details.

Position	Bioenact.	REC	ATC	Act. Inf.
Representations	non-rep.	culture	language	yes
Normativity	yes	yes	no	yes
Teleology	yes	no	no	no
Autonomy	meaningful	ur-intentional	mechanistic	formalizable
Computability	no	no	physical	yes

Before moving to the next section, it is important to emphasize how cognitive science is dominated by computational ideas, including approaches that are openly anti-computationalist. This, we believe, is certainly not a problem in itself, especially considering how it has provided rigorous support to research which is particularly abstract in nature. In this sense, what we will suggest in the following section is that, instead of aiming to deny or discard computation as an explanatory framework, we should seriously examine whether this framework is enough to encompass every material phenomenon that could play a role for life and mind.

4. On a Hypothetical Non-Algorithmic Component

4.1. Pervasiveness of Computation

By this point, we have a better general idea of how an autonomous system can exhibit intelligence in purely mechanistic terms, which, given our context, leads us to a logical question: Is this what we think *the mind* is? Or are we expecting something *more*? If something else is missing, then what is it?

We may start from anti-computationalist ideas, stemming from autopoiesis and the wider framework of theoretical biology, as proposed in [73,85,95,109]. These ideas ruled out the classical computationalist paradigm because the subjective take of any cognitive system would make impossible a truthful representation of any kind. In fact, as noted by Dell [178], these systems are of the same kind as those described by Ashby [78]: thermodynamically open, while informationally closed. This led to the conclusion that autonomous systems have no inputs or outputs and that living beings could not process information in a computational fashion. Nevertheless, while some classical premises were rejected, this does not really entail that their processes cannot be formalized computationally (see Abramova and Villalobos [139], Villalobos and Dewhurst [179] for a more detailed explanation).

This is better illustrated by the concept of enaction, whereby the interpretation a system performs is the consequence of system–environment coherence; the system has its own intrinsic logic, but its ongoing interaction with the environment imbues this logic with an external component which is not objectively truthful, but it is logically consistent. This consistency, in fact, provides a stepping stone toward contemporary views that endorse the notion of life as autonomous mechanisms, i.e., a form of blind intelligence, akin to evolution.

It is in this sense, and insofar as any form of consistency of this kind can be formalized as a set of consistent recursive regularities, realized as a physical structure that supports it (this is the role of embodied cognition in enactivism), that these processes can be described and conceptualized as modelable systems or even as physical computation [6,170,180,181].

As explained by Villalobos and Dewhurst [6], Dewhurst and Villalobos [182], computation, in this incredibly general sense, does not necessarily involve representation; hence, it is possible that the intelligence exhibited by cognitive systems can be understood as computational, or at least it is not theoretically opposed to notions like autonomy [42,181]. This rather general idea of computation is somehow related to the notions of mechanical process and effective method, which were developed in mathematical logic in the 1930s and out of which grew the more specific notions of computation that we are familiar with today [22,27,31]. Along these lines, computation would be the mathematical label or description for consistent mechanisms and processes that appear to be multiply realizable, probably including cognitive ones.

This seems bolstered by mounting evidence that computational processes can be realized by chemical reactions in the absence of living or even autonomous systems [7,183,184] and the fact that non-living physical systems are capable of exhibiting behaviors of the kind that we associate with life [5,185], thus hinting that intelligent behavioral patterns precede and are *subsumed by*—or maybe better, *forged in*—living systems [9].

Furthermore, this is also supported by cases where robotic models display proper responses to the point that they can help us predict the behavior of the animals they are modeled after [101,103,104,186], reaffirming the point that computation (insofar as mathematical descriptions and algorithmic implementations) seems to capture a fundamental mechanistic dimension of life, as also posited by autopoietic theories (including the FEP). All of this, within a broader framework oriented toward an understanding of the principles of life and cognition in computational or multiple realizable terms [43,56,84,170,187–189], has configured a new computationalist approach against other positions that consider life to be a special and irreplaceable substrate for the mind [4,88,119,159].

This, of course, does not mean that living systems are computers operating with the same logic as, for instance, a von Neumann architecture but that there is a logic underlying their ongoing changes, even if this logic is fundamentally intrinsic to them. The main point that is important to emphasize is that what we understand as intelligence has to be consistent to some degree so that the *mappings* that produce some behavior (perceptual, motor, or of any kind) will produce it reliably; otherwise, any form of coherence would be lost at the exact moment the process ends. This is, essentially, the function that natural numbers serve for common automata and the idea behind Godel's numbering in the first place (see Appendix A); continuous recursion over elements require that those elements be stable enough to produce the same state transitions, given the same states of the system and inputs.

The crux of the issue is that whatever the model for some cognitive function, since all models are abstractions of regularities underlying some observable phenomenon, then by definition, it will be restricted to computational descriptions in this rather general sense, firstly due to our understanding of reality in terms of causal relations and secondly by the very nature of the mechanisms that cognitive systems require.

We can consider the following example. A Turing machine, or any automaton for that matter, follows instructions from a state table; however, as we know, living beings do not to have a state table of that sort. Autonomous systems do not need a state table because their structure is in itself an embodied encoding of the *instructions* for behavior, and these instructions are distributed through all the parts and states of the physical system. After all, the *metaphoric* state table is a way of saying that there will be a consistent mechanical

progression of states depending on the symbols in the tape (the tape in this case would be the incoming signals from the environment), which is the case for autonomous systems.

Essentially, the central notion at stake is that of functional recursive mechanisms, insofar as a consistent procedure, for if there were no mechanisms to determine the changes of a system (its behavior), then there would not be a way for that system to enact consistent responses, leading to random or arbitrary changes and presumably to a fast entropic decay; intrinsic coherence and organization require consistency to begin with.

Thus, while we need not expect living systems to carry out computations in the same way as the model that formalizes them, the premise is that the sort of adaptive, intelligent behavior we associate with life must be cemented over a biological generation or instantiation of (cognitive) regularities that can be described by recursive formal languages and theory of computation.

In short, because life requires an order that implies consistent mechanisms, and because algorithms are descriptions of syntactic operations that formalize this notion of mechanism, anything non-algorithmic (not describable in terms of Turing-machines) would not, at least in theory, be suitable for the coherence needed for life and cognition.

Is this, however, really the end of the matter?

4.2. The Limits of Computational Intelligence

When we say that chemical operations can instantiate computations, what we are actually implying is that they are prone to computational descriptions. Insofar as every physical process that can be described in computational terms represents a regular and consistent sequence or network of events, these chemical transformations illustrate the link between the notions of (physical) mechanism and (mathematical) computation [7,183]. From this perspective, it is not strange to think that the bases of intelligence are materially *hardcoded* in physico-chemical properties.

Although, unlike living systems, the chemical solutions presented in Dueñas-Díez and Perez-Mercader [7] quickly dissipate, the point remains that regular, consistent patterns of change in a material universe can be abstracted and formalized as mathematical mechanistic regularities.

In this sense, what computational formalizations of autonomous dynamics describe are regularities that can be used to *automatize* behavior but not because a given system is really running a program (in a representationally and algorithmically realist sense); rather, autonomous behavior is essentially an amalgam of multiple elements undergoing state transitions while following mechanical causation. This gives rise to reliably adaptive responses, allowing self-preservation and creating regularities that we can idealize as input–output relations, even if these multiple parallel transitions occur at different domains and timescales.

Given that the structural transformations involved in even minimal intelligent behavior demand consistency, as the behavioral mechanisms producing appropriate responses would otherwise not be possible, then what differentiates intelligent systems would be the particular logic underpinning these consistent processes, without which they would not exhibit systematic behavior.

This is the same as saying that (1) a system in the same identical state, when paired with an identical state of its environment, will unavoidably produce an identical transformation [22,27,42] and (2) if such a system is also operationally closed, then the transformations that it undergoes will make the system persist—to be the *same system itself*—until the point when mechanical causation will produce its disintegration.

The importance of this fact for life is evident when we consider that, essentially, living beings are entities that preserve their organization through order. Order, in this sense,

cannot be dissociated from computation because it arises from a fundamental coherence, which underpins the beings' internal dynamics and their modulation of environmental perturbations. In this sense, every consistent response (i.e., structural transformation) that living beings undergo is the result of a never-ending—at least until their death—sequence of operations sustaining their existence as an organized unit [56,91,190].

In summary, the problem is the following. At one end of the spectrum, cognitivism portrays cognition as syntactic machines, seemingly leaving no conceptual space for meaning or a mind [12]. At the other end, approaches stemming from theoretical biology and autopoiesis, in spite of their explicit ambition to look for a phenomenological dimension, lead to fundamentally the same picture [57,137,191], with complex mechanistic amalgams which seem to be no more than physical realizations of abstract computational principles and are therefore void of a mind.

That from both ends of this conceptual space we reach the same conclusion leaves us in an uneasy position.

In this context, what we suggest is that a potential solution to this conundrum may come from exploration of the hypothetical non-algorithmic properties of life or mind. It is true that physical computability, insofar as consistent regularities allowing order and mechanistic processes, are a fundamental asset of biological systems. But this does not necessarily entail that everything that a living system is and that all that it does can be circumscribed within the computational domain. (This is a well-known idea thanks to Penrose [14] but already proposed by Rosen [47,48,49] and further developed by several others [11,12,15,50,51,53,54].)

In fact, it has recently been demonstrated that cultured (biological) neuronal networks on a multi-electrode array can perform information processing tasks such as copying and transferring information to other networks [192]. At the functional level, the information processing task can be described in terms that are computable. But at the level of the underlying mechanism performing the processing, it is quite far from obvious that the same can be said, especially given the inherent noisiness and complexity of the neuronal networks.

Similarly, there are a number of examples of physically realized systems performing tasks that are normally thought of as amenable to computational approaches but whose underpinning mechanisms seem outside the realm of computability. The use of an evolutionary search algorithm to develop extremely compact and efficient analog electronic controllers for autonomous mobile robots is an interesting case in point [193,194].

Garvie et al. [194] gave the first demonstration of transistor-level analog electronic controllers evolved directly in hardware for non-trivial, visually guided robot behaviors. This work highlighted how unconventional dynamics can be exploited for sensorimotor control by evolving controllers in a physical medium, thus tapping directly into its spatiotemporal and physical properties. A configurable transistor array chip was used to evolve the individual transistor properties (width and length, which determine the current and voltage characteristics) and the way in which the transistors were connected together and to the robot inputs and outputs. A series of careful experiments and analyses demonstrated that the controllers used indirect capacitive and field effects dependent on the particular physical peculiarities of the chip they were evolved on, so much so that moving the evolved circuit to another chip or another part of the same chip meant its performance degraded significantly.

Attempts to characterize the details of the underlying mechanisms at play proved to be quite difficult. Algorithmic, computable equivalents, derived from detailed analysis, failed to produce the same behavior, as did simplified equivalent circuits. The controllers were exploiting subtle, complex dynamics dependent on their physical implementation substrate, down to a deep—possibly molecular—level. It is difficult to imagine how these

dynamics can be abstracted away from the substrate and shown to be computable or an instance of algorithmic computation. It is possible that the only meaningful explanations possible will require a full account of the detailed physics of the substrate. Whether or not similar physics-dependent effects exist in biological nervous systems is an open question.

The remaining question then is whether there may be non-computational factors at play which could somehow propitiate biological and cognitive properties based not only upon computable properties. The answer is not straightforward, as too great a disruption would unavoidably produce the collapse of the cognitive system, which requires functionally stable mechanisms to dynamically persist and display intelligent behavior. Thus, even if we may lean toward the idea of a hypothetical non-computational component being useful for cognition, then this could not surpass a presumably incredibly small threshold beyond which the systems risks disintegration.

In this sense, it is important to depart from where we stand as of today, namely the idea that the correct architecture and enough complexity will eventually render ensembles of computable mechanisms into minded entities or at least a simulation of them. In this case, cognition and everything else would be the same thing, and therefore trees, computers, and brains would just be different types of instantiation of the same kind.

The alternative, or the idea that biological and mental organisms may somehow exploit some non computational feature of the physical world (with life and mind as blurry separations of diverse kinds), would need to be validated indirectly by characterizing or providing evidence of biological or cognitive phenomena that explicitly violate the former assumption.

4.3. Cognition and Protocognition—Intelligence, Life, and Mind

In our view, whatever life is, it is something beyond mere autonomous dynamics. Therefore, it is appropriate to label the specific kind of intelligent behavior organisms exhibit as cognitive behavior, inasmuch as the processes they realize surpass far simpler system–environment coherence and consistent interpretation-response mappings [2,3,56,142,148,159]. It does not follow from this, however, that every organism can be said to have a mind (insofar as sentient and purposeful control over their own states).

Similarly, to attribute intelligence only to living or mental beings is excessively arbitrary. Moreover, it seems to us that, as many others have hinted or explicitly pointed out [5,7,9,70,185], this may be better understood as a mechanistic property arising spontaneously from natural laws that leads to the formation of more intricate physico-chemical ensembles, eventually leading to autonomous self-referent systems and only much later to living systems as we know them.

A simple proof for this is the vast literature on artificial life and bio-inspired robotics and artificial intelligence, which has been able to emulate minimal cognitive-like behavior without the need for an organic *motherboard*. Such applications in robotics are especially illustrative [98,99,101,104,110,195,196].

Few would doubt that the mechanisms underlying the behavior of the systems presented in these works give rise to intelligent behavioral patterns. But whether this kind of intelligence is equivalent to that of biological systems (i.e., cognitive) is a different matter. In fact, we believe that the answer is no, because the different nature of the *hardware* will unavoidably lead to different computational specifications. In simple words, even if, to some extent, they may be solving the same problem—let us say navigation—they are posed differently; functionally speaking, the computational problem they specify is not entirely the same.

Thus, we suggest distinguishing as proto-cognitive properties the specific properties by which autonomous systems display intelligence and, by autonomy, a self-referential

property at its core [91,93,95]. Therefore, this is a particular kind of intelligent behavior which is determined by organizationally closed dynamics and which is non-mental and non-organic. Think, for instance, of behavior exhibited by inorganic chemical objects [7,185,197] or by artificial implementations, such as patterns in the Game of Life [10,87,198,199]. Put differently, a kind of intelligence that is logically and evolutionarily prior to life, while still constrained by a recursive nature, is such that the coherences that a system exhibits are not just transient or evanescent processes but safeguarded by being encoded in the structure of the entity that remains: a minimal autonomous system (see Figure 3 for a simple illustration).

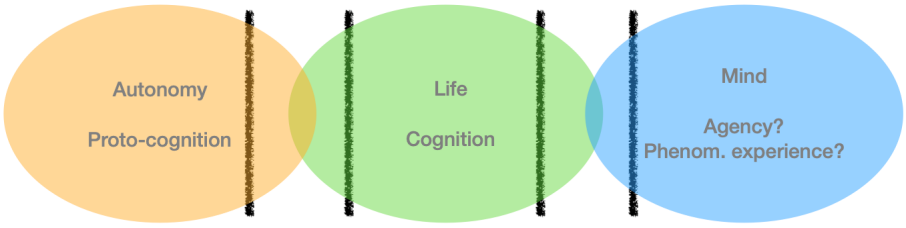

Figure 3. A simple schematic representing the proposed conceptual separation between proto-cognition, cognition, and mind. The three circles represent different stages of intelligence, proto-cognition (inorganic autonomous systems), cognition (the kind of intelligence specific to life), and mind (the one that it can be said to be the object of contemporary cognitive science). The space between the black lines separating the circles from each other represents a nonbinary cut from one category to the next. See the main text for further details.

Similarly, we believe that to conflate cognition, insofar as the specific kind of intelligence of living beings, with mind in the sense of phenomenological experience, agency, and other phenomena of this kind, would lead us to the same potential mistakes. In this sense, it seems somehow to be a wish to endow life with meaning or to bestow a phenomenological soul to living beings from the valuation we as humans make from our wonder as we confront it.

To put it simply, on the one hand, it seems evident that from the arbitrary set of entities to which we could attribute consciousness, aside from us, the more likely case is that of other living beings (a dog or a whale, for example). This, of course, is only intuition, even if extremely strong nonetheless.

On the other hand, we have no real proof whatsoever. Why living beings? Actually, when we think of organisms such as sunflowers, any single cell composing our organs, bacteria, or amoebae, to name a few, we suddenly realize that attributing sentience to every form of life may be precipitated. As a matter of fact, consciousness science is still too recent to be able to provide any definitive answers yet [4,175,177].

Being thus, regarding the hypothetical relation between cognition and phenomenological experience, if by the latter we understand someone or something that somehow experiences the very fact of its existence, even if minimally, then we believe that, given the current available knowledge, it is not only prudent but also sensible to remain agnostic.

In the next section, we will discuss agency as a possible case of non-computability, a feature that could, at least in theory, apply to cognitive and mental systems.

4.4. The Case of Agency

A different way to see this problem is by focusing on the hypothetical notion of agency, which reflects quite well this conflict between computable mechanisms and what we understand as a mind.

Intuitively, agency can be portrayed as the capacity of a given system to determine, at least to some extent, its own behavioral trajectory (to "decide" what to "do") or, in other

words, to display some degree of causal decoupling from the world that both surrounds it and realizes it. *This is something fundamentally real to our experience while, at the same time, fundamentally incompatible with a view in which every cognitive process is computable.*

Agency seems to somehow be a natural solution to a type of problem for which more archaic forms of adaptive behavior (such as pure structural adaptivity or reinforcement learning) are insufficient or do not fit as well. These problems include "simple" things, like reading a book or driving a car. This stage-like leap is appealing because agency can be conceived as a higher degree of intelligence, stemming from more complex cognitive structures. Unfortunately though, beyond intuitions of this kind, our actual understanding of the underlying mechanisms supporting any real form of agency are still quite unclear and have been a matter of increasing debate [106,149]. Roughly speaking, the main problem is that if every cognitive process is Turing-computable and therefore simulable and mechanistically describable in causal terms, then how could the *mind* of the system-agent disrupt the ongoing concatenation of causal events? Put differently, if every cognitive process is computable, then our experience of agency must be false.

In fact, lately there has been a proliferation of theoretical approaches concerned with the logical principles of agency, which have switched the spotlight toward the fundamental properties enabling or endowing a system with it [1,53,54,155,200,201]. This has resulted not only in a heated debate about the very notion of agency itself but has also exposed a conceptual frailty of related relevant notions. In particular, if everything is determined by (computable) causal mechanisms, then can we still talk about agency in any meaningful sense of the term?

If we go all the way back to Turing's [17] seminal paper on computable numbers, then a central difference is explicitly made between *automatic* or *a* machines, and *choice* or *c* machines (those that will require an *external* operator to eventually make choices for them). Turing Machines are *a* machines.

Along these lines, an autopoietic system is indeed an idealized organic version of an *a* machine, fully inter-constrained by the relational needs of its elements, namely an *autopoietic ouroboros*, something that applies to any model of autonomy as well. In this sense, the same underlying properties that permit the appearance of order, lawful behavior, and organized systems are precisely what preclude the possibility of agency, as we intuitively understand it at least.

Let us unpack this a bit. Protocognitive properties can be understood as those that produce intelligent behavior, which can be characterized through recursive mechanisms. In conceptual terms, these properties can be associated with autonomous systems. Mathematically, they can be described through theory of computation.

In our view, protocognition can explain phenomena such as goal-directed behavior, inasmuch as it is a mechanistically specified kind of behavior stemming from self-preserving dynamics. From some non-living systems such as droplets [185,197] or robotic implementations [99,100] all the way up to complex animals, goal-directed behavior would be the resulting activity of the system, arising from the myriad of interactions among its elements. And even if this may be incredibly complex, emergent, and nonlinear across multiple timescales, the fact remains the same; there is no conceptual space for *something* or *someone* to actually make a decision.

As we have seen through this essay, different concepts such as ultrastability [66], autopoiesis, autonomy [73,116,179], and ur-intentionality [137] examine this same problem from different perspectives. Moreover, empirical evidence from bio-inspired robotics and artificial life has shown that there is neither the need for a mind for collective intelligence to appear nor for a mind to guide the behavior of an intelligent system [106,202]. Something that we could say is that it was already well expressed by Leibniz's mill argument [203].

Furthermore, goal-directed behavior in living beings may be the most complex case of all, given the degree of amalgamation and subordination that it requires. In fact, it may be the limit of what can be characterized strictly through computational descriptions, insofar as an ensemble of blind mechanisms causally specifies the behavior of the unit as a whole without the need for meaning or a unified cognitive perspective. This is why different approaches speak of different, non-intersecting 'levels' [204,205]. Morphological computation and others, in this sense, by ascribing to an intentional stance fundamentally acknowledge cognition as a form of agency, even if due to practical reasons.

Here, the difference between goal-directed behavior and agency can be quite illustrative; whereas the former is fundamentally mechanistic (*a* machine describable), the latter involves and requires a mind that does not seem to fit these descriptions.

Insofar as any kind of cognitive mechanism can be represented computationally as a sequence of operations on a Turing machine, there will not be a conceptual space for agency. Indeed, this is a direct consequence of the Church–Turing thesis. Actually, to look for a hypothetical step or sequence of operations that could somehow subvert the cognitive concatenation of interdependent operations, cannot be anything other than a misconception; we will not find agency within computable notions like autonomy or free energy, however complex and intricate they be.

In other words, if we assume the existence of an agent that could induce (in the sense of some weak decision making) changes in its environment, then such a system must do so by producing changes in itself. This, however, would require the capacity of the system to somehow *decide* its own future states, which would require either for the system to be causally independent (even if infinitesimally briefly) or for another system that could act within it, which would lead to an infinite regression.

Even if we were to consider the notion of structural determinism [88,178], which is consistent with current theories on emergence [206,207], the problems remains very much the same, because the concatenation among processes leaves no 'gap' for the system to exploit underdetermined circumstances. The same applies to the notion of interactional asymmetry [1], which while conceptually right is limited by the same constraints in the absence of a phenomenological mind. If we think that the fundamental reason for this is sense making, a phenomenological property that cannot be reduced to computational descriptions, then the quest becomes how to characterize phenomenological phenomena through non-algorithmic models.

Sometimes, the problem is addressed from the lens of higher-level cognitive capacities, specifically those of a human most of the time. In this context, an interesting idea is temporal thickness (from predictive models) [112,169,208], whereby a system would generate models of future events while keeping a record of the events that just transpired. While the depth or extension of the temporal window generated will depend on the complexity of the system, the core premise could be understood as producing a cognitive capacity for evaluating possible futures.

The issue is that when we think of the brain as an agent in this way, that being as the source of the instructions that animate the body, we often forget to consider that the brain itself would be the mechanistic system and thus still bounded by computable state transition dynamics as a whole. Hence, if we would like to think that agency is maybe the product of the pre-frontal cortex or whatever other neural region or combination of them, then that system becomes the autonomous system bounded by the regularities that underpin its operation.

Therefore, and exactly as with the enactivist notion of agency, even if conceptually sound, there is no reason to believe that temporal thickness by itself could allow for a decision-making gap. Actually, this seems to lead in the totally opposite direction. Put

differently, whereas a complex enough system could, in many different ways, generate hypotheses about upcoming events in terms of probabilistic likelihood and define consistent responses in relation to its beliefs and expectations, all of this machinery is cognitive, thus being computational and, therefore, representable in terms of syntactic operations which are by definition devoid of mind or a mental subject, which we would have expected to be the (domain of the) agent.

In this sense, we suggest that agency may not be straightforwardly sub- or super-Turing [209], and that pressure to choose between them is part of the problem itself. On the one hand, sub-Turing would entail that organic intelligence be *weaker* than a physically instantiated approximation of a Turing machine. On the other hand, super-Turing would imply that biological systems somehow encode and *compute* undecidable problems, oracle machines, and similar matters, which is probably even more implausible than the former case. More importantly, both labels presuppose that cognition is, fully and fundamentally, some form of computable process.

Taking all this into consideration, we believe that it may be possible that living beings or some kind of system previous or after them, at some evolutionary stage, may have benefited from a non-computational component. Put differently, it may be the case that it is a sort of cognitive exploitation of the disruption of purely mechanistic processes, which marks a difference between pure mechanistic behavior and something closer to what we understand as a real *decision-making* agent.

Turing himself was open to the idea that there could be physical phenomena beyond what is computable [19], something that has also been suggested by others in the context of life and cognition (see, for example, Penrose [14], Rosen [47], Longo [51], Froese [53]). While most people seem to be skeptical about this, the truth is that we do not really know, and so to discard the possibility without any evidence may prove to be a misstep.

5. Summary and Conclusions

The fact that every general form of computation has resulted to be equivalent has led to the conjecture that any other will also be equivalent in expressive, so computational power. This conjecture, applied to any and every form of computation that could be, is therefore the claim of the Church-Turing thesis; that any possible form of computation—presumably including the ones that give rise to our thoughts and experiences—will be equivalent in computational power, and therefore modelable and simulable by others.

Following the Church–Turing thesis, every operation can be abstracted through the notion of cognitive mechanism [34], that is, through a form of consistent regularity (or a collection of them) by which a system specifies a response and which gives rise to its observable behavior. Certainly, given that the degree of selectivity of every system is different and dynamic, the environmental and internal changes that can be transduced as signals are limited and prone to error; however, cognitive systems are intrinsically consistent inasmuch as they interpret physical events mechanistically [116,182,210].

Thus, insofar as a problem can be well defined in algorithmic terms, there will be some syntactic machine capable of tackling it. Hence, *if* natural systems are to be conceived of as decision-making entities, their decisions should be characterizable as series of locally logical events (even if incredibly complex) and therefore formalized in terms of some abstract analog device representing their (environmental) inputs, internal states, and state transitions. In this sense, what we suggest is that eventually artificial life and other disciplines will have to seriously undertake the quest for an answer, for which the research question will presumably be an inversion of mainstream focus. That is, the problem to be solved will presumably not be the characterization of cognitive or Turing analogous models

but ways to "escape" from this. An example of this incipient process can be exemplified by the increasing interest in the discussion about agency [53,200,201,205,211].

Conceptually, there are two positions that we often take when examining the relation between life and computation: that living systems are capable of intelligent behavior because they instantiate computational properties or that we merely describe such intelligent behavior through the lens of mathematics and computation, but ultimately, the nature of life is something inherently different from such descriptions.

Along these lines, there are two related ideas that we should keep in mind. First, insofar as mathematical constructs (including that of computation) may be used to describe observable phenomena, there is no principled reason to either assert or discard, purely on the same mathematical grounds, the existence of non-computational processes concerning biological organizations. Secondly, given the existence of biological systems as natural systems, it is perfectly reasonable to at least consider the possibility that some of the current conceptual gaps may be a consequence of constraining our formal descriptions of biological phenomena exclusively to the scope of algorithmic models [11,15,52].

In other words, although we may acknowledge that everything that is protocognitive is formalizable by computational (i.e., mechanistic) means, that does not imply that sentient organisms may be constrained by the limits of our descriptions [14,50,51,53]. Thus, even if each one of the underlying regularities of cognitive systems were combinations of protocognitive features (and hence prone to computational descriptions), the case may always be that non-algorithmic phenomena have a small or even minimal influence, albeit strong enough to destabilize the otherwise imperturbable chain of (cognitive) events [27,51].

From this point of view, the life-mind-computation problem, originally a cognitive science dilemma, greatly overlaps with the core research motivation of artificial life [106,107,212,213], insofar as an exploration of the limits of purely mechanistic (autonomous) systems logically prior to life and the investigation of intelligent properties that could involve a non-computational component, such as sentience, agency, or temporality [53,106,155,164,172,201,214,215]. Along these same lines, efforts related to open-endedness and unconventional computing may prove fruitful in the long run [216–220].

Furthermore, the idea of exploitation of (non-algorithmic) phenomena through cognitive mechanisms provides a hypothetical solution for this contradiction and may provide insights toward a principled account of agency, similar to discussions in previous work on the possible roles of chaos or randomness in behavior generation [27,45,46,221]. In this sense, materiality may be seen as a mathematical, conceptual, or even a methodological conundrum, rather than a physical one.

On the other hand, to say that there are non-algorithmic properties of life would presumably entail a strong re-examination of some important assumptions in cognitive science, and so it should be approached even more carefully. Indeed, unlike deterministic chaos, undecidability has not been demonstrated to exist outside mathematics (i.e., in the physical world) [19,27,46], in spite of ideas such as randomness being tacitly agreed as true. A question that follows from this is whether some hypothetical form of biological undecidability would necessarily be underpinned by physics, or if it could stem from life as such, therefore being a different, non-reducible kind with its own explanatory level.

In this paper, we critically examined the idea of the Church–Turing thesis as the limit not only of what is computable but also of what is cognitive (problem explained in Section 2). We made several conceptual clarifications to show how deeply connected cognition and computation are and to portray the problematic search for a *mind* (Section 3). Finally, we have proposed that current limits in cognitive science may relate to limiting cognition to computation, and we illustrated the issue through the case of agency (Section 4).

To some extent, the work we have presented is comparative, insofar as different ways to approach the relation between cognition and computation are portrayed and compared in Section 3, but also analytical, insofar as we attempted to break down the problem and focus on one issue, namely that computation, if understood in the context of a physical realization of recursive and consistent mechanisms underpinning intelligent behavior, cannot explain all cognitive phenomena.

The idea we developed throughout this paper is a theoretical conjecture and not a metaphysical postulate (i.e., the possible biological and cognitive exploitation of some non-computable physical phenomenon). Furthermore, we have shown how originally and supposedly anti-computationalist frameworks are largely algorithmic themselves. The main issue, we reckon, relates to formalization, inasmuch as "proper" understanding in science requires neat reproducibility.

We are trained in science to formalize and categorize (otherwise we have not "understood"), which almost inevitably leads to algorithmic theories of cognition. Perhaps biological cognition cannot be constrained in this way. What we suggest is that it would be wrong to dismiss non-algorithmic possibilities, even if we may not be able to formalize them in the traditional scientific way. Further work in this direction would have to be particularly aware and careful of this.

Author Contributions: Conceptualization, F.R.-V. and P.H.; validation, P.H.; investigation, F.R.-V.; writing—original draft preparation, F.R.-V.; writing—review and editing, F.R.-V. and P.H.; visualization, F.R.-V.; supervision, P.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study.

Acknowledgments: The authors would like to especially thank Adam Rostowski for his insights into these problems and his collaboration with the conference version of this paper, as well as the anonymous reviewers, from ALife and for this version, who provided thoughtful comments and advice.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Recursion in Godel's Numbering

In order to introduce the notion of a recursive language, we need to review in more detail the relation between recursion and computation. For this, we will briefly illustrate the idea by referring to Godel's numbering (for a more detailed explanation, see Raatikainen [28]).

First, we will associate the primitive symbols s of a formal language F to arbitrary natural numbers $\#(s)$ such that a non-comprehensive list would be:

$$\#(\text{`0'}) = 1$$

$$\#(\text{`}\prime\text{'}) = 2$$

$$\#(\text{`+'}) = 3$$

$$\#(\text{`}\times\text{'}) = 4$$

$$\#(\text{`=`}) = 5$$

Then, based on the unique prime factorization theorem, by which we know that any natural number greater than one can be obtained from the product of the powers of prime numbers, we can represent any number x with

$$x = p_1^{n_1} \times p_2^{n_2} \times ... \times p_k^{n_k}$$

where the subindices stand for the prime numbers starting from one until the kth prime number (i.e., $2, 3, 5, ..., p_k$) and the superindices $(n_1, n_2, ..., n_k)$ for the correspondent powers of each of these prime numbers. Given that the sequence of prime numbers is known, we can represent this simply with the notation $\langle n_1, n_2, ..., n_k \rangle$ determining the powers. The following is a minimal example:

$$x = \langle 5, 1, 3 \rangle = 2^5 \times 3^1 \times 5^3 = 32 \times 3 \times 125 = 12{,}000$$

Then, by combining both these features (encoding over natural numbers plus prime factorization), it is possible to assign a symbol number to any expression derived from the formal rules of F. These encodings of symbolic numbers as unique associations between a sequence of numbers and a formal expression are Godel numbers. Again, as a simple example, let us consider the following case:

$$1 \times 0 = 0$$

Given that natural numbers can be represented by using successors (one is the successor of zero, two is the successor of the successor of zero, etc.), from the associations above $\#(\text{`0'}) = 1, \quad \#(\text{`}\prime\text{'}) = 2$ (which represents the successor of), and $\#(\text{`}\times\text{'}) = 4$, the Godel number for this expression will be given by

$$\langle 1, 2, 4, 1, 5, 1 \rangle = 2^1 \times 3^2 \times 5^4 \times 7^1 \times 11^5 \times 13^1 = 164{,}875{,}961{,}250$$

Hence, the expression $\ulcorner 1 \times 0 = 0 \urcorner$ is uniquely encoded by the Godel number 164,875,961,250 within the formal system F. Conversely, there is a unique mechanical procedure for the translation of natural numbers into their respective formulas and derivations (even if many numbers do not actually encode one), which is obtained from the inverse procedure. Moreover, whereas we have presented a quite trivial example, recursion by means of these symbolic identifiers allows for the definition of more complex syntactic properties and mechanical operations, such as negation, variables, implication, proofs, and formal derivations [28].

As a matter of fact, Godel used this mathematical construction to prove that (1) for every formal system F with a minimal arithmetical complexity, there will be statements that cannot be proven or disproven and (2) that the consistency of any F cannot be proven within the same F. These proofs are generally known as the *incompleteness theorems* and were the first concrete case exhibiting the limits of computation, for which Turing's halting problem can be seen as a generalization (see Perales-Eceiza et al. [27] for a historical account).

Appendix B. Comments on conceptual foundations

1. While the terms computable, recursive, and decidable are roughly interchangeable these days, in theory, they are to be applied to specific mathematical structures, namely sets, functions, and formal languages, respectively.
2. Note that the fact that something is effectively computable does not necessarily imply that it can be efficiently computed, that is, within a *reasonable* amount of time. After all, a Turing machine is an abstract concept with virtually infinite time to find a solution. Indeed, it may be effectively calculable, but over exponential increases in the expected running time. In this sense, depending on the length of the inputs, it is therefore intractable in *realistic* terms. The field of study of this kind of problem is commonly referred to as computational complexity theory.

3. Simply put, recursion permits the definition of recursive mathematical subsets for which an operationally closed language is determined. Recalling the more basic notion of computable sets, if the function f (a Turing machine in this case) is a total function such that it is defined on every element of the domain B from which the domain of the formal language A is a subset, then the definition of recursive is adhered to because the Turing machine halts for every input. In these cases, the Turing machine is sometimes called a total Turing machine and is said to be the *decider* of the recursive language, as it is able to decide and recognize what strings belong to the language and which do not. In this same vein, note that decidability and computability refer fundamentally to the same issue, but a problem is said to be decidable if the output (the answer to the problem) can be posed in binary terms, whereas computability refers to all the cases surpassing this one-bit output (yes- or no-like answer).
4. While we can simply take this as a convergent series with a finite total sum (i.e., $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + ... = 1$), what Zeno is illustrating is the difference between mathematical and material realms. Simply put, in a physical universe, it is impossible to infinitely subdivide space in halves because at some point, one arrives at an indivisible unit (the Planck length, as far as we know); therefore, even if incredibly large, the number of parts or distances is not infinite, and thus it can be realized in finite time.
5. Interestingly enough, all contemporary computational implementations such as neural networks, quantum, or neuromorphic computing systems, among many others, are still bounded within the capacity of Turing machines. So much so that in most cases they can be better associated to some kind of finite state automata (even if incredibly powerful ones in some cases). What is more, there seems to be no clear limit to the sort of physical realizations that we could deem as *computational*. In principle at least, every kind of different physical phenomena could be associated with some mechanistic description corresponding to some domain in the theory of computation, such as in the Chomsky hierarchy [19,42], to the point that it has been suggested that the whole material universe may be another equivalent instantiation of the same underlying principles (i.e., that the universe itself may be a specific kind of Turing machine, presumably a cellular automaton) [19,27,44] (We reckon, however, that it is better to leave these controversial although interesting ideas to science fiction for now [222]). On the flip side, an alternative view is defended and actively pursued by research on hypercomputation—hypothetical models of computation more powerful than Turing machines and therefore beyond the capabilities of the Church–Turing thesis. Although Turing himself rejected the possibility of *oracle* machines [223], there has been a late resurgence of the idea. See Copeland [222], Syropoulos [224] for more information.

Appendix C. Glossary of Enactive Related Notions

We have included this rather small glossary to disambiguate some terms that may be slightly confusing for the reader. Unfortunately, there are no universally accepted definitions for some of these terms. Precisely because of this, we have included Section 3, aiming to provide the reader with a clear overall picture. For further details on this issue, we recommend some more specific examinations referenced in the main text, such as the following:

- Boden, 2000 [225];
- Abramova and Villalobos, 2015 [139];
- Ward, Silverman, and Villalobos, 2017 [114];
- Barandiaran, 2017 [96];
- Villalobos and Silverman, 2018 [57];

- Gallagher, 2018 [55];
- Gallagher, 2023 [115].

► **Autopoiesis** (based on Maturana and Varela [73])

An autopoietic system is a network of processes of production (transformation and destruction) of components capable of the following:

1. It regenerates and realizes the network of processes that produced them;
2. It constitutes the system as a concrete unity in space by specifying its own boundary.

In other words, let the following be true:

- $P = \{P_i\}$ is a set of production processes;
- $C = \{C_i\}$ is a set of components.

Then, the system is autopoietic *iff* the following are true:

- $P \rightarrow C$ and $C \rightarrow P$ form a closed production network;
- There is a boundary B where the following are true:
  - i B is produced by the network such that $B \subset C$;
  - ii B confines the material realization of C to its internal domain.

► **Autonomy** (based on Varela [91])

Varela [91] is explicit about the following points:

1. Autonomy is organizational, not functional or representational;
2. Identity precedes interaction (identity is not defined by inputs or outputs);
3. Closure is operational, not energetic or informational;
4. Autonomy is graded, not binary (unlike autopoiesis).

Let the following be true:

- $P = \{P_i\}$ is a set of processes;
- $R \subseteq PxP$ is an enabling relation between processes;
- $S = \{S_i\}$ is the set of valid states of the unitary or global system.

Then, the system S is autonomous *iff* the following are true:

- S realizes an organizational network $\mathcal{O} = (P, R)$;
- $\mathcal{O} = (P, R)$ is operationally closed (see the next glossary entry);
- $\mathcal{O} = (P, R)$ is an invariant class such that $\mathcal{O}_t \simeq \mathcal{O}_{t+i}$;
- The organization is invariant to a class of environmental perturbations E such that the following are true:
  1. $\forall S_i, S_j \in S$ is a state transition $S_i \rightarrow S_j \Rightarrow \mathcal{O}(S_i); \simeq \mathcal{O}(S_j)$;
  2. $\mathcal{O}_t \simeq \mathcal{O}_{t+\Delta t}$ after any perturbation $e \in E$.

Finally, given that autonomy is graded rather than binary, the degree to which the system is autonomous will depend on how much the system processes determine its own structural transformations such that

$$A(S) = \frac{R_S}{R_{S,E}}$$

See [94,199,226] for further discussion and examples.

► **Operational closure** (based on Varela [91])

Operational closure (also known as organizational closure) is the central axiom presented in [91], underpinning the notion of autonomy. Following from the definitions from the previous entry, a system is operationally closed *iff* the following are true:

- $\forall p_i \in P$, $\exists\{p_{j1}, p_{j2}, ..., p_{jn}\} \subseteq P$ such that the following are true:
  1. These can stablish enabling relations, where $(p_i, p_j) \in R$;
  2. These relations are enabled only by $\forall(p_i, p_j) \in R$, where $p_i, p_j \in P$.

In simple words, processes enable other processes through enabling relations that only enable processes that belong to the organizational set of processes (i.e., mediating system structural states). Therefore enabling relations cannot leave the organizational domain, because processes can only enable other processes by means of the same kind of enabling relations, and the operation of the system closes in itself.

► **Adaptivity** (based on Di Paolo et al. [1], Di Paolo [124])

The concept of adaptivity departs from the following:

1. Autonomy and operational closure (both are assumed as correct);
2. Viability conditions;
3. Normative evaluation of the system states relative to those conditions;
4. Endogenous modulation of behavior (based on a normative evaluation).

Hence, let the following be true:

- $x(t) \in \mathcal{X}$ is the state of the system at time t within the state-space $\mathcal{X}$;
- $V \subset \mathcal{X}$ is the set of viable states;
- ∂V is the boundary of viability;
- $v : \mathcal{X} \to R$, where the following are true:
  - $v(x) > 0 \Rightarrow x \in V$ (within the viable domain);
  - $v(x) = 0 \Rightarrow x \in \partial V$ (approaching "breakdown" conditions);
  - $v(x) < 0 \Rightarrow \notin V$ (out of the viable domain; disintegration).
- $\frac{d}{dt}v(x(t))$ is an internal regulatory capacity to modulate the system dynamics.

This is such that an autonomous system is adaptive *iff* the following are true:

1. $\exists p \in P$ such that $\dot{x} = f(x, v(x))$;
2. $v(x) \to 0 \Rightarrow \mathcal{F}[\frac{d}{dt}v(x)] > 0$.

This is to say that (1) there are internal dynamics that vary as a function of the viability of the system ($v(x)$), and (2) regulation is normative, because approaching non-viability ($v(x) = 0$) triggers dynamics that counter this tendency and increase viability, where $\mathcal{F}$ represents a tendency in the dynamics of the system under perturbation (toward viability in this case).
Given that the mechanisms regulating the dynamics are internally generated and stemming from the operational closure of the system, they are physically constrained and graded. This can be expressed as

$$Adp(S) = \mathcal{F}\,[\frac{d}{dt}v(x) \,|\, v(z) \simeq 0]$$

where higher values of $Adp(S)$ indicate higher adaptive capacities.

► **Cognitive mechanism**

Any form of consistent regularity (or a collection of them) by which a system specifies a response stemming from its own states and environmental perturbations and which gives rise to its observable behavior. Let the following be true:

- $\{Si\} = S$ is the set of states representing valid material configuration of the system;
- $\{e_i\} = E$ is the set of environmental perturbations modulating the system.

Then, every cognitive mechanism will produce changes of the following forms:

1. Given $\mathcal{O} : (S_x, e) \rightarrow S_y$, where $S_x, S_y \in S$ and $e \in E$;
2. $\nabla S_x \neq \nabla S_y$;
3. $\mathcal{O}(S_x^t) \simeq \mathcal{O}(S_y^{t+\Delta t})$.

Hence, whereby (1) the system transitions into other valid states with respect to environmental circumstances (i.e., it adapts), (2) the selectivity and the corresponding sensor and behavioral capacities of the system (represented by ∇) will vary among states while (3) preserving its organization.

References

1. Di Paolo, E.; Burhmann, T.; Barandarian, X. *Sensorimotor Life: An Enactive Proposal*; Oxford University Press: Oxford, UK, 2017.
2. Lyon, P.; Cheng, K. Basal cognition: Shifting the center of gravity (again). *Anim. Cogn.* **2023**, *26*, 1743–1750. [CrossRef]
3. Lyon, P.; Keijzer, F.; Arendt, D.; Levin, M. Reframing cognition: Getting down to biological basics. *Phil. Trans. R. Soc. B* **2021**, *376*, 20190750. [CrossRef]
4. Seth, A. Conscious artificial intelligence and biological naturalism. *PsyArXiv* **2024**. [CrossRef]
5. McGregor, S.; Virgo, N. Life and its Close Relatives. In *Advances in Artificial Life. Darwin Meets von Neumann. ECAL 2009*; Lecture Notes in Computer Science; Kampis, G., Karsai, I., Szathmáry, E., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; Volume 5778.
6. Villalobos, M.; Dewhurst, J. Why post-cognitivism does not (necessarily) entail anti- computationalism. *Adapt. Behav.* **2017**, *25*, 117–128. [CrossRef]
7. Dueñas-Díez, M.; Perez-Mercader, J. How Chemistry Computes: Language Recognition by Non-Biochemical Chemical Automata. From Finite Automata to Turing Machines. *iScience* **2019**, *19*, 514–526. [CrossRef]
8. Villalobos, M.; Palacios, S. Autopoietic theory, enactivism, and their incommensurable marks of the cognitive. *Synthese* **2021**, *198*, 571–587. [CrossRef]
9. Egbert, M.; Hanczyc, M.M.; Harvey, I.; Virgo, N.; Parke, E.C.; Froese, T.; Sayama8, H.; Penn, A.S.; Bartlett, S. Behaviour and the Origin of Organisms. *Orig. Life Evol. Biosph.* **2023**, *53*, 87–112. [CrossRef] [PubMed]
10. Agüera y Arcas, B.; Alakuijala, J.; Evans, J.; Laurie, B.; Mordvintsev, A.; Niklasson, E.; Randazzo, E.; Versari, L. Computational Life: How Well-formed, Self-replicating Programs Emerge from Simple Interaction. *arXiv* **2024**, arXiv:2406.19108v2.
11. Rubin, S. Cartography of the multiple formal systems of molecular autopoiesis: From the biology of cognition and enaction to anticipation and active inference. *BioSystems* **2023**, *230*, 104955. [CrossRef]
12. Searle, J.R. Minds, brains, and programs. *Behav. Brain Sci.* **1980**, *3*, 417–457. [CrossRef]
13. Dreyfus, H.L. *What Computers Still Can't Do: A Critique of Artificial Reason*; MIT Press: Cambridge, MA, USA, 1992.
14. Penrose, R. *Shadows of the Mind. A Search for the Missing Science of Consciousness*; Oxford University Press: Oxford, UK, 1994.
15. Froese, T.; Taguchi, S. The Problem of Meaning in AI and Robotics: Still with Us after All These Years. *Philosophies* **2019**, *4*, 14. [CrossRef]
16. Boyer, C.B. *A History of Mathematics*; John Wiley and Sons: Hoboken, NJ, USA, 1968.
17. Turing, A.M. On Computable Numbers, with an Application to the Entscheidungsproblem. *Proc. Lond. Math. Soc.* **1936**, *s2–42*, 230–265. [CrossRef]
18. Copeland, J. *Artificial Intelligence: A Philosophical Introduction*; Wiley-Blackwell: Hoboken, NJ, USA, 1993.
19. Copeland, J.; Shagrir, O. Is the whole universe a computer? In *The Turing Guide: Life, Work, Legacy*; Copeland, J., Bowen, J., Sprevak, M., Wilson, R., Eds.; Oxford University Press: Oxford, UK, 2017.
20. Church, A. An Unsolvable Problem of Elementary Number Theory. *Am. J. Math.* **1936**, *58*, 345–363. [CrossRef]
21. Church, A. A Note on the Entscheidungsproblem. *J. Symb. Log.* **1936**, *1*, 40–41. [CrossRef]
22. Turing, A.M. Computing Machinery and Intelligence. *Mind New Ser.* **1950**, *59*, 433–446. [CrossRef]
23. Hopcroft, J.E.; Ullman, J.D. *Formal Languages and Their Relation to Automata*; Addison-Wesley Publishing Company: Reading, MA, USA, 1969.
24. Church, A. Review of: On Computable Numbers with an Application to the Entscheidungsproblem by A.M. Turing. *J. Symb. Log.* **1943**, *2*, 42. [CrossRef]
25. Post, E.L. Recursively Enumerable Sets of Positive Integers and Their Decision Problems. *Bull. Am. Math. Soc.* **1944**, *50*, 284–316. [CrossRef]

26. Liesbeth, D. Turing Machines. In *The Stanford Encyclopedia of Philosophy (Summer 2025 Edition)*; Zalta, E.N., Nodelman, U., Eds.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2025. Available online: https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=turing-machine (accessed on 25 January 2026).
27. Perales-Eceiza, A.; Cubitt, T.; Gu, M.; Pérez-García, D.; Wolf, M.M. Undecidability in Physics: A Review. *Phys. Rep.* **2025**, *1138*, 1–29. [CrossRef]
28. Raatikainen, P. Gödel's Incompleteness Theorems. In *The Stanford Encyclopedia of Philosophy (Fall 2025 Edition)*; Zalta, E.N., Nodelman, U., Eds.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2025. Available online: https://plato.stanford.edu/entries/goedel-incompleteness/ (accessed on 25 January 2026).
29. Hofstadter, D.R. *Godel, Escher, Bach: An Eternal Golden Braid*; Basic Books: Hassocks, UK, 1979.
30. Huggett, N. Zeno's Paradoxes. In *The Stanford Encyclopedia of Philosophy (Fall 2024 Edition)*; Zalta, E.N., Nodelman, U., Eds.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2024. Available online: https://plato.stanford.edu/archives/win2025/entries/paradox-zeno/ (accessed on 25 January 2026).
31. Copeland, B. What is Computation. *Synthese* **1996**, *108*, 335–359. [CrossRef]
32. Van Gelder, T. What Might Cognition Be, If Not Computation? *J. Philos.* **1995**, *92*, 345–381. Available online: http://www.jstor.org/stable/2941061?origin=JSTOR-pdf (accessed on 25 January 2026). [CrossRef]
33. Baltieri, M.; Buckley, C.L.; Bruineberg, J. Predictions in the eye of the beholder: An active inference account of Watt governors. In Proceedings of the ALIFE2020: The 2020 Conference on Artificial Life, Online, 13–18 July 2020.
34. Bermudez, J.L. *Cognitive Science. An Introduction to the Science of the Mind*; Texas A & M University: College Station, TX, USA, 2022.
35. Copeland, B. The Church-Turing Thesis. *The Stanford Encyclopedia of Philosophy (Winter 2024 Edition)*; Zalta, E.N., Nodelman, U., Eds.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2024. Available online: https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=church-turing (accessed on 25 January 2026).
36. Kleene, S.C. λ-Definability and Recursiveness. *Duke Math. J.* **1936**, *2*, 340–353. [CrossRef]
37. Dean, W.; Naibo, A. Recursive Functions. In *The Stanford Encyclopedia of Philosophy (Summer 2025 Edition)*; Zalta, E.N., Nodelman, U., Eds.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2025. Available online: https://plato.stanford.edu/archives/sum2025/entries/recursive-functions/ (accessed on 25 January 2026).
38. Chomsky, N. On Certain Formal Properties of Grammars. *Inf. Control* **1959**, *2*, 137–167. [CrossRef]
39. Davis, M. *Computability & Unsolvability*; McGraw-Hill Book Company: New York, NY, USA, 1958.
40. Gardner, M. Mathematical Games: The Fantastic Combinations of John Conway's New Solitaire Game 'Life'. *Sci. Am.* **1970**, *223*, 120–123. [CrossRef]
41. Lewis, H.R.; Papadimitriou, C.H. *Elements of the Theory of Computation*; Prentice-Hall: Upper Saddle River, NJ, USA, 1981.
42. Piccinini, G.; Corey, M. Computation in Physical Systems. In *The Stanford Encyclopedia of Philosophy (Summer 2021 Edition)*; Zalta, E.N., Nodelman, U., Eds.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2021. Available online: https://plato.stanford.edu/archives/fall2025/entries/computation-physicalsystems/ (accessed on 25 January 2026).
43. Goni-Moreno, A. Biocomputation: Moving Beyond Turing with Living Cellular Computers. *Commun. ACM* **2024**, *67*, 70–77. [CrossRef]
44. Perales-Eceiza, A.; Cubitt, T.; Gu, M.; Pérez-García, D. Undecidability in Physics: A Review. *arXiv* **2024**, arXiv:2410.16532v1. Available online: https://arxiv.org/html/2410.16532v1#S1 (accessed on 25 January 2026).
45. Zenil, H. *Randomness Through Computation: Some Answers, More Questions*; World Scientific: Singapore, 2011.
46. Agüero, J.M.; Calude, T.S.; Dinneen, M.J.; Fedorov, A.; Kulikov, A.; Navarathna, R.; Svozil, K. How Real is Incomputability in Physics? *arXiv* **2024**. Available online: https://arxiv.org/pdf/2311.00908 (accessed on 25 January 2026).
47. Rosen, R. Some realizations of (M, R)-systems and their interpretation. *Bull. Math. Biophys.* **1971**, *33*, 303–319. [CrossRef] [PubMed]
48. Rosen, R. Some realizations cell models: The metabolism-repair systems. In *Foundations of Mathematical Biology. Cellular Systems*; Academic Press: Cambridge, MA, USA, 1972; pp. 217–273.
49. Rosen, R. *Life Itself: A Comprehensive Inquiry Into the Nature, Origin, and Fabrication of Life*; Columbia University Press: New York, NY, USA, 1991.
50. Louie, A.H.; Poli, R. The spread of hierarchical cycles. *Int. J. Gen. Syst.* **2011**, *40*, 237–261. [CrossRef]
51. Longo, G. From exact sciences to life phenomena: Following Schrödinger and Turing on Programs, Life and Causality. *Inf. Comput.* **2009**, *207*, 545–558. [CrossRef]
52. Lane, P.A. Robert Rosen's Relational Biology Theory and His Emphasis on Non-Algorithmic Approaches to Living Systems. *Mathematics* **2024**, *12*, 3529. [CrossRef]
53. Froese, T. Irruption Theory: A Novel Conceptualization of the Enactive Account of Motivated Activity. *Entropy* **2023**, *25*, 748. [CrossRef]
54. Jaeger, J.; Vervaeke, J.; Riedl, A.; Djedovic, A.; Walsh, D. Naturalizing relevance realization: Why agency and cognition are fundamentally not computational. *Front. Psychol.* **2024**, *15*, 1362658. [CrossRef]

55. Gallagher, S. 4E Cognition. Historical Roots, Key Concepts, and Central Issues. In *The Handbook of 4E Cognition*; Newen, A., de Bruin, L., Gallagher, S., Eds.; Oxford University Press: Oxford, UK, 2018.
56. Allen, M.; Friston, K. From cognitivism to autopoiesis: Towards a computational framework for the embodied mind. *Synthese* **2018**, *195*, 2459–2482. [CrossRef]
57. Villalobos, M.; Silverman, D. Extended functionalism, radical enactivism and the autopoietic theory of cognition: Prospects for a full revolution in cognitive science. *Phenomenol. Cogn. Sci.* **2018**, *17*, 719–739. [CrossRef]
58. Boden, M.A. Life and Mind. *Minds Mach.* **2009**, *19*, 453–463. [CrossRef]
59. Villalobos, M. Enactive cognitive science: Revisionism or revolution? *Adapt. Behav.* **2013**, *21*, 159–167. [CrossRef]
60. Di Paolo, E.A.; Thompson, E.; Beer, R.D. Laying down a forking path: Tensions between enaction and the free energy principle. *Philos. Mind Sci.* **2022**, *3*, 1–39. [CrossRef]
61. Miller, G.A. The cognitive revolution: A historical perspective. *Trends Cogn. Sci.* **2003**, *7*, 141–144. [CrossRef]
62. Tolman, E. Cognitive maps in rats and men. *Psychol. Rev.* **1948**, *55*, 189–208. [CrossRef] [PubMed]
63. Miller, G.A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.* **1956**, *63*, 81–97. [CrossRef]
64. Rosenblatt, F. The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **1958**, *65*, 386–408. [CrossRef]
65. McCulloch, W.; Pitts, W. A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bull. Math. Biophys.* **1943**, *7*, 115–133. [CrossRef]
66. Ashby, W. *Design for a Brain: The Origin of Adaptive Behavior*; Chapman and Hall: Boca Raton, FL, USA, 1960.
67. Chomsky, N. *Syntactic Structures*; Mouton: Gravenhage, The Netherlands, 1957.
68. Minsky, M. A framework for representing knowledge. *Artif. Intell.* **1974**, *Memo 306*, 1–81. Reprinted in *The Psychology of Computer Vision*; Winston, P., Ed.; McGraw-Hill: Columbus, OH, USA, 1975.
69. Fodor, J. *The Language of Thought*; Harvard University Press: Cambridge, MA, USA, 1975.
70. Newell, A.; Simon, H.A. Computer Science asEmpirical Inquiry: Symbols and Search. *Commun. ACM* **1976**, *19*, 113–126. [CrossRef]
71. Marr, D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*; W. H. Freeman: San Francisco, CA, USA, 1982.
72. Newell, A. Physical Symbol Systems. *Cogn. Sci.* **1980**, *4*, 135–183. [CrossRef]
73. Maturana, H.; Varela, F. *Autopoiesis: The Organization of the Living. [De maquinas y seres vivos. Autopoiesis: La organizacion de lo vivo]*, 7th ed. from 1994; Editorial Universitaria: Santiago, Chile, 1973.
74. Rosenblueth, A.; Wiener, N.; Bigelow, J. Behavior, Purpose and Teleology. *Philos. Sci.* **1943**, *10*, 18–24. [CrossRef]
75. Wiener, N. *Control and Communication in the Animal and the Machine*; Wiley and Sons: New York, NY, USA, 1948.
76. von Foerster, H. *Understanding Understanding: Essays on Cybernetics and Cognition*; Springer: Berlin/Heidelberg, Germany, 2013.
77. Walter, W. An Imitation of Life. *Sci. Am.* **1950**, *1825*, 42–45. [CrossRef]
78. Ashby, W. *An Introduction to Cybernetics*; J. Wiley: New York, NY, USA, 1956.
79. Lettvin, J.Y.; Maturana, H.R.; McCulloch, W.S.; Pitts, W.H. What the Frog's Eye Tell the Frog's Brain. *Proc. IRE* **1959**, *47*, 1940–1951. [CrossRef]
80. Varela, F.; Maturana, H.; Uribe, R. Autopoiesis: The organization of the living systems, its characterization and a model. *Biosystems* **1974**, *5*, 187–196. [CrossRef]
81. Letelier, J.; Marin, G.; Mpodozis, J. Autopoietic and (M,R) systems. *J. Theor. Biol.* **2003**, *22*, 261–272. [CrossRef]
82. Virgo, N.; Biehl, M.; Baltieri, M.; Capucci, M. A "good regulator theorem" for embodied agents. In Proceedings of the ALIFE 2025: Proceedings of the 2025 Artificial Life Conference, Copenhagen, Denmark, 14–18 July 2025.
83. Ashby, W. Principles of the self-organizing dynamic system. *J. Gen. Psychol.* **1947**, *37*, 125–128. [CrossRef]
84. Razeto-Barry, P. Autopoiesis 40 years Later. A Review and a Reformulation. *Orig. Life Evol. Biosph.* **2012**, *42*, 543–567. [CrossRef]
85. Maturana, H.; Varela, F. *The Tree of Knowledge: The Biological Roots of Human Understanding*; New Science Library/Shambhala Publications: Boulder, CO, USA, 1987.
86. Beer, R. Autopoiesis and Cognition in the Game of Life. *Artif. Life* **2004**, *10*, 309–326. [CrossRef]
87. Beer, R. The Cognitive Domain of Glider in the Game of Life. *Artif. Life* **2014**, *20*, 183–206. [CrossRef]
88. Maturana, H. Autopoiesis, Structural Coupling and Cognition: A history of these and othe notions in the biology of cognition. *Cybern. Hum. Knowing* **2002**, *9*, 5–34.
89. Lyon, P. Autopoiesis and Knowing: Reflections on Maturana's Biogenic Explanation of Cognition. *Cybern. Hum. Knowing* **2004**, *11*, 21–46.
90. Maturana, H.R. Reality: The Search for Objectivity or the Quest for a Compelling Argument. *Ir. J. Psychol.* **1988**, *9*, 25–82. [CrossRef]
91. Varela, F. *Principles of Biological Autonomy*; North Holland: Amsterdam, The Netherlands, 1979.

92. Varela, F.J. The Creative Circle: Sketches on the Natural Hisotry of Circularity. In *The Invented Reality*; Watzlavick, P., Ed.; Norton Publishing: New York, NY, USA, 1984.
93. Varela, F. Two Principles for Self-Organization. In *Self-Organization and Management of Social Systems*; Ulrich, H., Probst, G.J.B., Eds.; Springer Series on Synergetics; Springer: Berlin/Heidelberg, Germany, 1984; Volume 26.
94. Varela, F. Structural Coupling and the Origin of Meaning in a Simple Cellular Automation. In *The Semiotics of Cellular Communication in the Immune System*; Sercarz, E.E., Celada, F., Mitchison, N.A., Tada, T., Eds.; NATO ASI Series; Springer: Berlin/Heidelberg, Germany, 1988; Volume 23.
95. Varela, F.; Thompson, E.; Rosch, E. *The Embodied Mind: Cognitive Science and Human Experience*; The MIT Press: Cambridge, MA, USA, 1991.
96. Barandiaran, X. Autonomy and Enactivism: Towards a Theory of Sensorimotor Autonomous Agency. *Topoi* **2017**, *36*, 409–430. [CrossRef]
97. Hinton, G.E.; Anderson, J.A. *Parallel Models of Associative Memory*; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1989.
98. Brooks, R. Intelligence without representation. *Artif. Intell.* **1991**, *47*, 139–159. [CrossRef]
99. Cliff, D.; Husbands, P.; Harvey, I. Explorations in Evolutionary Robotics. *Adapt. Behav.* **1993**, *2*, 73–110. [CrossRef]
100. Beer, R. A dynamical systems perspective on agent-environment interaction. *Artif. Intell.* **1995**, *72*, 173–215. [CrossRef]
101. Beer, R. The dynamics of adaptive behavior: A research program. *Artif. Intell.* **1997**, *20*, 257–289. [CrossRef]
102. Harvey, I.; Di Paolo, E.; Wood, R.; Quinn, M. Evolutionary Robotics: A New Scientific Tool for Studying Cognition. *Artif. Life* **2005**, *11*, 79–98. [CrossRef]
103. Webb, B. Robots in invertebrate neuroscience. *Nature* **2002**, *417*, 359–363. [CrossRef]
104. Baddeley, B.; Graham, P.; Husbands, P.; Philippides, A. A Model of Ant Route Navigation Driven by Scene Familiarity. *PLoS Comput. Biol.* **2012**, *8*, e1002336. [CrossRef]
105. Gershenson, C. Requisite Variety, Autopoiesis, and Self-organization. *Kybernetes* **2015**, *44*, 866–873. [CrossRef]
106. Baltieri, M.; Iizuka, H.; Witkowsi, O.; Sinapayen, L.; Suzuki, K. Hybrid Life: Integrating biological, artificial, and cognitive systems. *Wires Cogn. Sci.* **2023**, *14*, e1662. [CrossRef]
107. Aguilar, W.; Santamaria-Bonfil, G.; Froese, T.; Gershenson, C. The past, present and future of artificial life. *Front. Robot. AI* **2014**, *1*, 8. [CrossRef]
108. Husbands, P.; Holland, O.; Wheeler, M. (Eds.) *The Mechanical Mind in History*; MIT Press: Cambridge, MA, USA, 2008.
109. Maturana, H. Biology of cognition. In *Biological Computer Laboratory, BCL Report 9*; University of Illinois: Urbana-Champaign, IL, USA, 1970.
110. Tani, J. *Exploring Robotics Minds. Actions, Symbols, and Consciousness as Self-Organizing Dynamic Phenomena*; Oxford University Press: Oxford, UK, 2017.
111. Dennett, D. Autonomy, Consciousness, and Freedom. *Amherst Lect. Philos.* **2019**, *14*, 1–22.
112. Friston, K. Am I Self-Conscious? (Or Does Self-Organization Entail Self-Consciousness?). *Front. Psychol.* **2018**, *9*, 1–10. [CrossRef] [PubMed]
113. Bowes, S. *Naturally Minded: Mental Causation, Virtual Machines, and Maps*; Springer: Berlin/Heidelberg, Germany, 2023.
114. Ward, M.; Silverman, D.; Villalobos, M. Introduction: The Varieties of Enactivism. *Topoi* **2017**, *36*, 365–375. [CrossRef]
115. Gallagher, S. *Embodied and Enactive Approaches to Cognition*; Cambridge University Press: Cambridge, UK, 2023.
116. Varela, F. Patterns of life: Intertwining identity and cognition. *Brain Cogn.* **1997**, *34*, 72–87. [CrossRef]
117. Weber, A.; Varela, F. Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenol. Cogn. Sci.* **2002**, *1*, 97–125. [CrossRef]
118. Thompson, E. Sensorimotor subjectivity and the enactive approach to experience. *Phenomenol. Cogn. Sci.* **2005**, *4*, 407–427. [CrossRef]
119. Thompson, E. *Mind in Life: Biology, Phenomenology and the Sciences of Mind*; Harvard University Press: Cambridge, MA, USA, 2007.
120. Nave, K. Every Body's Gotta Eat: Why Autonomous Systems Can't Live on Prediction-Error Minimization Alone. Ph.D. Thesis, University of Edinburgh, Edinburgh, UK, 2022.
121. Rostowski, A. Freedom: An enactive possibility. *Hum. Aff.* **2022**, *32*, 427–438. [CrossRef]
122. Merleau-Ponty, M. *Phenomenology of Perception*; Routledge: London, UK, 2012.
123. Jonas, H. *The Phenomenon of Life: Toward a Philosophical Biology*; Northwestern University Press: Evanston, IL, USA, 1966.
124. Di Paolo, E. Autopoiesis, adaptivity, teleology, agency. *Phenomenol. Cogn. Sci.* **2005**, *4*, 429–452. [CrossRef]
125. Villalobos, M.; Ward, D. lived experience and cognitive science. Reappraising enactivism's Jonasian turn. *Constr. Found.* **2016**, *11*, 802–831.
126. Varela, F. Organism: A Meshwork of Selfless Selves. In *Organism and the Origins of Self. Boston Studies in the Philosophy of Science*; Tauber, A.I., Ed.; Springer: Dordrech, The Netherlands, 1991; Volume 129. [CrossRef]
127. Stewart, J. Cognition = Life: Implications for higher-level cognition. *Behav. Process.* **1996**, *35*, 311–326. [CrossRef]

128. Kirchhoff, M.; Froese, T. Where There is Life There is Mind: In Support of a Strong Life-Mind Continuity Thesis. *Entropy* **2017**, *19*, 169. [CrossRef]
129. Di Paolo, E. Organismically-inspired robotics: Homeostatic adaptation and natural teleology beyond the closed sensorimotor loop. In *Dynamical Systems Approach to Embodiment and Sociality*; Murase, K., Asakura, T., Eds.; Advanced Knowledge International: Adelaide, Australia, 2003.
130. Froese, T. To Understand the Origin of Life We Must First Understand the Role of Normativity. *Biosemiotics* **2021**, *14*, 657–663. [CrossRef]
131. Barandiaran, X.; Di Paolo, E.; Rohde, M. Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-temporality in Action. *Adapt. Behav.* **2009**, *17*, 367–386. [CrossRef]
132. Buhrmann, T.; Di Paolo, E. The sense of agency—A phenomenological consequence of enacting sensorimotor schemes. *Phenomenol. Cogn. Sci.* **2017**, *16*, 207–236. [CrossRef]
133. Barandiaran, X.; Moreno, A. Adaptivity: From Metabolism to Behavior. *Adapt. Behav.* **2008**, *16*, 325–344. [CrossRef]
134. Beer, R.; Di Paolo, E. The theoretical foundations of enaction: Precariousness. *Biosystems* **2023**, *223*. [CrossRef]
135. Hutto, D.D. Knowing what? Radical versus conservative enactivism. *Phenomenol. Cogn. Sci.* **2005**, *4*, 389–405. [CrossRef]
136. Hutto, D.D.; Myin, E. Deflating deflationism about mental represen- tation. In *What Are Mental Representations?* Smortchkova, J., Dołrega, K., Schlicht, T., Eds.; Oxford University Press: Oxford, UK, 2020.
137. Hutto, D.; Myin, E. *Evolving Enactivism. Basic Minds Meet Content*; MIT Press: Cambridge, MA, USA, 2017.
138. Rowlands, M. Hard Problems of Intentionality. *Philosophia* **2015**, *43*, 741–746. [CrossRef]
139. Abramova, K.; Villalobos, M. The apparent (Ur-)Intentionality of Living Beings and the Game of Content. *Philosophia* **2015**, *43*, 651–668. [CrossRef]
140. Maturana, H., Preface from Humberto Maturana Romesîn to the second edition. In *De Mâquinas y seres vivos. Autopoiesis: La organizaciôn de lo vivo*; Editorial Universitaria: Santiago, Chile, 1994; Chapter Preface.
141. Villalobos, M. Autopoiesis, Life, Mind and Cognition: Bases for a Proper Naturalistic Continuity. *Biosemiotics* **2013**, *6*, 379–391. [CrossRef]
142. Lyon, P. The biogenic approach to cognition. *Cogn. Process.* **2006**, *7*, 11–29. [CrossRef]
143. Wasserman, E.A. Comparative Cognition: Beginning the Second Century of the Study of Animal Intelligence. *Psychol. Bull.* **1993**, *113*, 211–228. [CrossRef]
144. Beran, M.; Parrish, A.; Perdue, B.; Agnes, S. Comparative Cognition: Past, Present, and Future. *Int. J. Comp. Psychol.* **2014**, *27*, 3–30. [CrossRef]
145. Shettleworth, S.J. *Fundamentals of Comparative Cognition*; Oxford University Press: Oxford, UK, 2012.
146. Baluška, F.; Levin, M. On Having No Head: Cognition throughout Biological Systems. *Front. Psychol.* **2016**, *7*, 1–19. [CrossRef] [PubMed]
147. Keijzer, F.; van Duijn, M.; Lyon, P. What nervous systems do: Early evolution, input-output, and the skin brain thesis. *Adapt. Behav.* **2013**, *21*, 1–19. [CrossRef]
148. Levin, M.; Keijzer, F.; Lyon, P.; Arendt, D. Uncovering cognitive similarities and differences, conservation and innovation. *Phil. Trans. R. Soc. B* **2021**, *376*, 20200458. [CrossRef] [PubMed]
149. Levin, M.; Dennet, D.C. Cognition all the way down. *Aeon* **2020**. Available online: https://aeon.co/essays/how-to-understand-cells-tissues-and-organisms-as-agents-with-agendas (accessed on 25 January 2026).
150. Adamatzky, A. *Physarum Machines. Computers from Slime Mould*; World Scientific Series on Nonlinear Science Series A; World Scientific: Singapore, 2010; Volume 74.
151. Vallverdú, J.; Castro, O.; Mayne, R.; Talanov, M.; Levin, M.; Baluška, F.; Gunji, Y.; Dussutour, A.; Zenil, H.; Adamatzky, A. Slime mould: The fundamental mechanisms of biological cognition. *Biosystems* **2018**, *165*, 57–70. [CrossRef]
152. Blackiston, D.; Lederer, E.; Kriegman, S.; Garnier, S.; Bongard, J.; Levin, M. A cellular platform for the development of synthetic living machines. *Sci. Robot.* **2021**, *6*, eabf1571. [CrossRef]
153. Levin, M. Bioelectric networks: The cognitive glue enabling evolutionary scaling from physiology to mind. *Anim. Cogn.* **2023**, *26*, 1865–1891. [CrossRef]
154. Dennet, D.C. The intentional stance in theory and practice. In *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*; Byrne, R.W., Whiten, A., Eds.; Clarendon Press/Oxford University Press: Oxford, UK, 1988.
155. Seifert, G.; Sealander, A.; Marzen, S.; Levin, M. From reinforcement learning to agency: Frameworks for understanding basal cognition. *BioSystems* **2024**, *235*, 105107. [CrossRef]
156. Friston, K. The history of the future of the Bayesian brain. *NeuroImage* **2012**, *62*, 1230–1233. [CrossRef]
157. Clark, A. Whatever Next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **2013**, *36*. [CrossRef]
158. Clark, A. A nice surprise? Predictive processing and the active pursuit of novelty. *Phenomenol. Cogn. Sci.* **2018**, *17*, 521–534. [CrossRef]

159. Wiese, W.; Friston, K. Examining the Continuity between Life and Mind: Is There a Continuity between Autopoietic Intentionality and Representationality? *Philosophies* **2021**, *6*, 18. [CrossRef]
160. Pezzulo, G.; Parr, T.; Andy Clark, P.C.; Friston, K. Generating meaning: Active inference and the scope and limits of passive AI. *Trends Cogn. Sci.* **2024**, *28*, 97–112. [CrossRef]
161. Friston, K. The free-energy principle: A rough guide to the brain? *Trends Cogn. Sci.* **2009**, *13*, 293–301. [CrossRef] [PubMed]
162. Seth, A.; Tsakiris, M. Being a beast machine: The somatic basis of selfhood. *Trends Cogn. Sci.* **2018**, *22*, 969–981. [CrossRef] [PubMed]
163. Clark, A. *Surfing Uncertainty: Prediction, Action and the Embodied Mind*; Oxford University Press: Oxford, UK.
164. Varela, F. Present-time consciousness. *J. Conscious. Stud.* **1999**, *6*, 111–140.
165. Thompson, E.; Varela, F.J. Radical embodiment: Neural dynamics and consciousness. *Trends Cogn. Sci.* **2001**, *5*, 418–425. [CrossRef]
166. Gallagher, S. The Past, Present and Future of Time-Consciousness: From Husserl to Varela and Beyond. *Constr. Found.* **2017**, *13*, 91–97.
167. Varela, F.J. Neurophenomenology: A methodological remedy for the hard problem. *J. Conscious. Stud.* **1996**, *3*, 330–349.
168. Varela, F.J. The Specious Present: A Neurophenomenology of Time Consciousness. In *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*; Petitot, J., Varela, F.J., Pachoud, B., Roy, J.-M., Eds.; Stanford University Press: Stanford, CA, USA, 1999.
169. Roseboom, W.; Seth, A.K.; Sherman, M.T.; Fountas, Z. The Perception of Time in Humans, Brains, and Machines. *PsyArXiv* **2022**, *2022* . [CrossRef]
170. Korbak, T. Computational enactivism under the free energy principle. *Synthese* **2021**, *198*, 2743–2763. [CrossRef]
171. Albarracin, M.; Pitliya, R.J.; Ramstead, M.J.D.; Yoshimi, J. Mapping Husserlian phenomenology onto active inference. *arXiv* **2022**, arXiv:2208.09058. [CrossRef]
172. Bogotá, J.D.; Djebbara, Z. Time-consciousness in computational phenomenology: A temporal analysis of active inference. *Neurosci. Conscious.* **2023**, *2023*, niad004. [CrossRef] [PubMed]
173. Ramstead, M.J. Naturalizing what? Varieties of naturalism and transcendental phenomenology. *Phenomenol. Cogn. Sci.* **2015**, *14*, 929–971. [CrossRef]
174. Ramstead, M.J.; Seth, A.K.; Hesp, C.; Sandved-Smith, L.; Mago, J.; Lifshitz, M.; Pagnoni, G.; Smith, R.; Dumas, G.; Lutz, A.; et al. From Generative Models to Generative Passages: A Computational Approach to (Neuro) Phenomenology. *Rev. Philos. Psychol.* **2022**, *13*, 829–857. [CrossRef]
175. Albertazzi, L. Naturalizing Phenomenology: A Must Have? *Front. Psychol.* **2018**, *9*, 1933. [CrossRef]
176. Koch, C.; Massimini, M.; Boly, M.; Tononi, G. Neural correlates of consciousness: Progress and problems. *Nat. Rev. Neurosci.* **2016**, *17*, 307–321. [CrossRef]
177. Gallagher, S. On the possibility of naturalizing phenomenology. In *The Oxford Handbook of Contemporary Phenomenology*; Zahavi, D., Ed.; Oxford Academic: Oxford, UK, 2012.
178. Dell, P. Understanding Bateson and Maturana: Toward a Biological Foundation for The Social Sciences. *J. Marital Fam. Ther.* **1985**, *11*, 1–20. [CrossRef]
179. Villalobos, M.; Dewhurst, J. Enactive autonomy in computational systems. *Synthese* **2018**, *195*, 1891–1908. [CrossRef]
180. Putnam, H. Minds and Machines. In *Dimensions of Mind: A Symposium*; Hook, S., Ed.; Collier: New York, NY, USA, 1960.
181. Piccinini, G.; Scarantino, A. Information processing, computation, and cognition. *J. Biol. Phys.* **2011**, *37*, 1–38. [CrossRef] [PubMed]
182. Dewhurst, J.; Villalobos, M. The Enactive Automaton as a Computing Mechanism. *Thought* **2017**, *6*, 185–192. [CrossRef]
183. Perez-Mercader, J.; Dueñas-Diez, M.; Case, D. Chemically-Operated Turing Machine, 2013. U.S. Patent 20140200716A1, 28 February 2017.
184. López-Díaz, A.J.; Sayama, H.; Gershenson, C. The Origin and Evolution of Information Handling. *arXiv* **2024**, arXiv:2404.04374. [CrossRef]
185. Hanczyc, M.; Ikegami, T. Chemical basis for minimal cognition. *Artif. Life* **2010**, *16*, 233–243. [CrossRef] [PubMed]
186. Webb, B. Can robots make good models of biological behaviour? *Behav. Brain Sci.* **2001**, *24*, 1033–1050. [CrossRef]
187. Agüera y Arcas, B.; Fairhall, A.L.; Bialek, W. Computation in a Single Neuron: Hodgkin and Huxley Revisited. *Neural Comput.* **2003**, *15*, 1715–1749. [CrossRef]
188. Sayama, H. Construction theory, self-replication, and the halting problem. *Complexity* **2008**, *13*, 16–22. [CrossRef]
189. Rouleau, N.; Levin, M. The Multiple Realizability of Sentience in Living Systems and Beyond. *Cogn. Behav.* **2023**, *10*, 1–7. [CrossRef] [PubMed]
190. Maturana, H. The organization of the living: A theory of the living organization. *Int. J. Man-Mach. Stud.* **1975**, *7*, 313–332. [CrossRef]
191. Tallis, R. *Freedom: An Impossible Reality*; Agenda Publishing Limited: London, UK, 2021.

192. Roberts, T.; Kern, F.; Fernando, C.; Szathmary, E.; Husbands, P.; Philippides, A.; Staras, K. Encoding Temporal Regularities and Information Copying in Hippocampal Circuits. *Sci. Rep.* **2019**, *9*, 19036. [CrossRef]
193. Husbands, P.; Shim, Y.; Garvie, M.; Dewar, A.; Domcsek, N.; Graham, P.; Nowotny, T.; Philippides, A. Recent advances in evolutionary and bio-inspired adaptive robotics: Exploiting embodied dynamics. *Appl. Intell.* **2021**, *51*, 6467–6496. [CrossRef]
194. Garvie, M.; Flascher, I.; Philippides, A.; Thompson, A.; Husbands, P. Evolved transistor array robot controllers. *Evol. Comput.* **2020**, *28*, 677–708. [CrossRef]
195. Quinn, M. Evolving Communication without Dedicated Communication Channels. In *Advances in Artificial Life. ECAL 2001*; Kelemen, J., Sosík, P., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2001; Volume 2159. [CrossRef]
196. Egbert, M.; Barandiaran, X. Modeling habits as self-sustaining patterns of sensorimotor behavior. *Front. Hum. Neurosci.* **2014**, *8*, 1–15. [CrossRef]
197. Löffler, R.J.G.; Hanczyc, M.M.; Gorecki, J. A hybrid camphor–camphene wax material for studies on self-propelled motion. *Phys. Chem. Chem. Phys.* **2019**, *21*, 24852–24856. [CrossRef] [PubMed]
198. Beer, R. Bittorio revisited: Structural coupling in the Game of Life. *Adapt. Behav.* **2020**, *28*, 197–212. [CrossRef]
199. Rodriguez-Vergara, F.; Husbands, P. Proto-Cognitive Bases of Agency. *Preprints* **2025**. [CrossRef]
200. Potter, H.; Mitchell, K. Naturalasing agent causation. *Entropy* **2022**, *24*, 472. [CrossRef]
201. Biehl, M.; Virgo, N. Interpreting systems as solving POMDPs: A step towards a formal understanding of agency. In *Active Inference. IWAI 2022*; Buckley, C.L., Cialfi, D., Lanillos, P., Ramstead, M., Sajid, N., Shimazaki, H., Verbelen, T., Eds.; Communications in Computer and Information Science; Springer: Cham, Switzerland, 2023; Volume 1721. [CrossRef]
202. Husbands, P. Never Mind the Iguana, What About the Tortoise? Models in Adaptive Behavior. *Adapt. Behav.* **2009**, *17*, 320–324. [CrossRef]
203. Duncan, S. Leibniz's Mill Arguments Against Materialism. *Philos. Q.* **2012**, *62*, 250–272. [CrossRef]
204. Maturana, H. Ultrastability... autopoiesis? Reflective response to Tom Froese and John Stewart. *Cybern. Hum. Knowing* **2011**, *18*, 143–152.
205. Virgo, N.; Biehl, M.; McGregor, S. Interpreting Dynamical Systems as Bayesian Reasoners. In *Machine Learning and Principles and Practice of Knowledge Discovery in Databases. ECML PKDD 2021*; Kamp, M., Koprinska, I., Bibal, A., Bouadi, T., Frénay, B., Galárraga, L., Oramas, J., Adilova, L., Krishnamurthy, Y., Kang, B., et al., Eds.; Communications in Computer and Information Science; Springer: Cham, Switzerland, 2021; Volume 1524. [CrossRef]
206. Mediano, P.A.; Rosas, F.; Carhart-Harris, R.L.; Seth, A.K.; Barrett, A.B. Beyond integrated information: A taxonomy of information dynamics phenomena. *arXiv* **2019**, arXiv:1909.02297. [CrossRef]
207. Varley, T.F. Flickering Emergences: The Question of Locality in Information-Theoretic Approaches to Emergence. *Entropy* **2022**, *25*. [CrossRef] [PubMed]
208. Seth, A. *Being You: A New Science of Consciousness*; Faber and Faber Ltd.: London, UK, 2021.
209. Douglas, K. *Super-Turing Computation: A Case Study Analysis*; Technical Report; Carnegie Mellon University: Pittsburgh, PA, USA, 2003.
210. Maturana, H.; Uribe, G.; Frenk, S. A biological theory of relativistic colour coding in the primate retina. *Arch. Biol. Med. Exp.* **1968**, *1*, 1–30.
211. Horiguchi, L.; Maruyama, N.; Shigeto, D.; Crosscombe, M.; Ikegami, T. Quantifying Autonomy in Ant Colonies Using Non-Trivial Information Closure. In Proceedings of the ALIFE 2024: Proceedings of the 2024 Artificial Life Conference, Copenhagen, Denmark, 22–26 July 2024. [CrossRef]
212. Boden, M.A. Creativity and ALife. *Artif. Life* **2015**, *21*, 354–365. [CrossRef]
213. Pigozzi, F. Of Typewriters and PCs. In Proceedings of the ALIFE 2023: Ghost in the Machine, Sapporo, Japan, 24–28 July 2023. [CrossRef]
214. Yamashita, Y.; Tani, J. Emergence of Functional Hierarchy in a Multiple Timescale Neural Network Model: A Humanoid Robot Experiment. *PLoS Comput. Biol.* **2008**, *4*, e1000220. [CrossRef]
215. Rodriguez, F.; Husbands, P.; Ghosh, A.; White, B. Frame by frame? A contrasting research framework for time experience. In Proceedings of the ALIFE 2023: Ghost in the Machine: Proceedings of the 2023 Artificial Life Conference, Sapporo, Japan, 24–28 July 2023; p. 75. [CrossRef]
216. Soros, L.; Stanley, K. Identifying necessary conditions for open-ended evolution through the artificial life of Chromaria. In Proceedings of the ALIFE 14: Proceedings of the Fourteenth International Conference on the Synthesis and Simulation of Living Systems, New York, NY, USA, 30 July–2 August 2014; pp. 793–800. [CrossRef]
217. Stepney, S. Modelling and measuring open-endedness. In Proceedings of the OEE4 Workshop, at ALife 2021, Prague, Czech Republic (Online), 19–23 July 2021.
218. Ackley, D.; Small, T. Indefinitely Scalable Computing = Artificial Life Engineering. In Proceedings of the ALIFE 14: The Fourteenth International Conference on the Synthesis and Simulation of Living Systems, New York, NY, USA, 30 July–2 August 2014.

219. Stepney, S. Programming Unconventional Computers: Dynamics, Development, Self-Reference. *Entropy* **2012**, *14*, 1939–1952. [CrossRef]
220. Broersma, H.; Stepney, S.; Wendin, G. Computability and Complexity of Unconventional Computing Devices. In *Computational Matter*; Stepney, S., Rasmussen, S., Amos, M., Eds.; Springer: Cham, Switzerland, 2019. [CrossRef]
221. Shim, Y.; Husbands, P. Embodied neuromechanical chaos through homeostatic regulation. *Chaos* **2019**, *29*, 033123. [CrossRef]
222. Copeland, B. Hypercomputation. *Minds Mach.* **2002**, *12*, 461–502. [CrossRef]
223. Turing, A. Systems of Logic Based on Ordinals. *Proc. Lond. Math. Soc.* **1939**, *s2*, 161–228. [CrossRef]
224. Syropoulos, A. *Hypercomputation. Computing Beyond the Church-Turing Barrier*; Springer Nature: Berlin/Heidelberg, Germany, 2008.
225. Boden, M.A. Autopoiesis and Life. *Cogn. Sci. Q.* **2000**, *1*, 117–145.
226. Bertschinger, N.; Olbrich, E.; Ay, N.; Jost, J. Autonomy: An information theoretic perspective. *BioSystems* **2008**, *91*, 331–345. [CrossRef]