Strange (mental) machines. About computation, protocognition and Artificial Life

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

[1]AI Research Group, University of Sussex, UK
[2]Sussex Centre for Consciousness Science, University of Sussex, UK
f.rodriguez-vergara@sussex.ac.uk

Abstract

Enactivism has had an important conceptual influence on Artificial Life. Its notorious mind-life continuity thesis is a fundamental idea that once sustained much of the theoretical framework, but has recently lost support, explicitly in cognitive science, and by omission in Artificial Life. Alternative computational accounts, however, do not seem capable of filling the vacuum it leaves behind, given that according to the Church-Turing thesis, the limit of what is computable is bounded by Turing machines (regardless of the specific instantiation, such as cellular automata, deep, spiking or any other form of neural networks, or even quantum computers, to name some well known cases) and that computable functions formally describe the notion of mechanism, every organismic process that determines a consistent response should be both limited to Turing-machine capabilities and amenable to formalization. This impending issue, namely, the relation among life, cognition and computation as the hypothetical elements of the mind, is particularly pertinent and highly relevant for Artificial Life. Moreover, although it is often overlooked or omitted for practical reasons, it will probably define, or so we argue, the direction of future research in cognitive science, artificial intelligence and related fields.

Introduction

For around half a century now, a biological approach to cognition has slowly but surely influenced our understanding of cognitive phenomena, shifting previous conceptions from a rather anthropocentric and representational view, to one where the mind appears deeply entangled with life (Varela et al., 1991; Boden, 2009; Di Paolo et al., 2017; Lyon and Cheng, 2023). This has instigated reconsideration of fundamental notions such as intelligence, behavior, computation, cognition and life.

This conceptual shift has influenced different areas of cognitive science and related disciplines, such as comparative cognition (Wasserman, 1993; Shettleworth, 2012; Beran et al., 2014; Levin et al., 2021), theoretical biology (Lyon, 2004, 2006; Keijzer, 2015; Levin, 2023), minimal/basal cognition (van Duijn et al., 2006; Garzon and Keijzer, 2011; Lyon and Cheng, 2023), bioinspired artificial systems and robotics (Brooks, 1991; Cliff et al., 1993; Beer, 1995; Webb, 2002; Harvey et al., 2005), and where the establishment of the field of Artificial Life (Aguilar et al., 2014; Pigozzi, 2023) can be seen as a crystallization of this process.

Although artificial life has consolidated itself as a field in its own right, it still bears strong conceptual ties to its theoretical mother branch: enactivism. (Varela, 1997; Gallagher, 2023). While there are plenty of discrepancies within the enactive framework (Barandiaran, 2017; Ward et al., 2017; Rubin, 2023) (some of which we will briefly revise in the following section) there are central concepts, like autopoiesis (Maturana and Varela, 1973; Stano et al., 2023; Gershenson et al., 2020) and autonomy (Maturana, 1975; Varela, 1979), which are conceptually related, to varying degrees, to all of them. Similarly, concepts like embodied, embedded cognition, enaction and sense-making were introduced to cognitive science by Varela et al. (1991), forming the foundations of enactivism's account of what it is to be a living system, of their very nature. (Varela, 1997; Thompson, 2007; Froese and Di Paolo, 2011; Hutto and Myin, 2017; Di Paolo et al., 2017). This is the main reason why, for enactivism, life and mind are intimately entangled.

The life-mind relation is not an exclusive concern of enactivism (Boden, 2009). Other approaches, such as minimal/basal cognition (Lyon, 2004, 2006; Garzon and Keijzer, 2011; Lyon et al., 2021; Lyon and Cheng, 2023; Levin, 2023), active inference (Clark, 2013; Seth and Tsakiris, 2018; Pezzulo et al., 2024), and the Free Energy Principle (Allen and Friston, 2018; Wiese and Friston, 2021; Bruineberga et al., 2022), have stemmed at varying degrees from the notion of autopoiesis (Maturana and Varela, 1973, 1987) (inasmuch as life is considered to be the source, or the domain of cognitive phenomena). The key difference, nonetheless, has to do with the degree of commitment to a sentient, phenomenological kind of mind, from which non-enactivist accounts usually prefer to remain more or less agnostic in the context of *basic* living systems (e.g. bacteria, or trees).

Furthermore, there has been a relatively recent theoretical push to reconsider some aspects of cognition in terms of pure *raw* mechanistic properties that might be implemented ubiquitously and independently of life. That is, as multiple realizable processes implemented by living organisms, with-

out being exclusive to them (McGregor and Virgo, 2011; Villalobos and Dewhurst, 2017; Dueñas-Díez and Perez-Mercader, 2019; Villalobos and Palacios, 2021; Egbert et al., 2023; Baltieri et al., 2023; Bowes, 2023; López-Díaz et al., 2024) – prefiguring a return to computationalist ideas in cognitive science (Newell and Simon, 1976; Newell, 1980).

Insofar as we understand cognitive properties as mechanisms underpinning intelligent capacities and/or behavior exhibited by living beings, while acknowledging that these mechanisms are consistent regularities that can be described computationally, the confrontation between computational and phenomenological aspects of life becomes unavoidable. Thus, given the continuous developments in artificial intelligence, the exploration of the theoretical relation between life and mind is evermore pressing. Indeed, Artificial Life will probably play a central role as an eventual arbiter to this dilemma.

In this context, we begin with a general review of one major fault line within enactive cognitive science, focusing on differences in views on the life-mind relation, before re-evaluating these divergences under different notions of computation, emphasizing the concept of (cognitive) mechanism, insofar as a consistent regularity that is necessary for all intelligent behavior.

Then, we examine the relation between cognition and protocognition, explaining the importance of delimiting between those properties exhibited by living beings (cognitive) and those which do not, but that enable intelligence while pre-existing life. These, therefore, are necessary, but not sufficient for cognition and thereby we conceive and refer to them as *proto-cognitive* properties. Finally, we present a brief summarizing discussion in the context of Artificial Life, to emphasize the importance of future research in this direction.

The Enactive conundrum

In a nutshell, the gap between (algorithmic) mechanisms and human experience has been a source of conflict in enactive cognitive science (Ward et al., 2017; Barandiaran, 2017; Villalobos and Palacios, 2021; Gallagher, 2023). As we will see in this section, this has fragmented the enactivist landscape into some different positions.

Bioenactivism

The attempt to reconnect formal developments in cognitive science with phenomenological insights, 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 (Varela, 1997; Weber and Varela, 2002; Thompson, 2005, 2007; Nave, 2022; Rostowski, 2022). Central to this turn is both the philosophical influence of phenomenological authors like Husserl, Merleau-Ponty (Merleau-Ponty, 2012), and especially from Jonas (1966), whose ideas set the basis for the life-mind continuity thesis (i.e., the idea that mind emerges from the living), and, on the introduction of new concepts and reformulations of problematic assumptions from the original ideas of autopoiesis (Di Paolo, 2005; Villalobos and Ward, 2016; Barandiaran, 2017), still very much part of the ideas in Varela et al. (1991).

Subscribers to stronger, or more traditional, versions of the life-mind continuity thesis hold that research in autonomy (in the strict enactive sense of a specific mode of operational closure inseparable from an embodied sensorimotor profile) forms a basis for understanding the emergence of an *experiencer* from subjectless systems (Varela, 1991; Stewart, 1996; Thompson, 2007; Barandiaran, 2017; Kirchhoff and Froese, 2017). Put another way, the constant threat of disintegration would bootstrap a need for and course of action into self-individuating systems, towards energy sources and 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 by the built-in imperative to keep existing, hence with teleological behavior (Weber and Varela, 2002; Di Paolo, 2003; Froese, 2021).

A fundamental guiding assumption for bioenactivism is that autonomous sensorimotor systems are *agents*, so capable of actively and asymmetrically determining their behavior (Barandiaran et al., 2009; Buhrmann and Di Paolo, 2017; Di Paolo et al., 2022) and, in this respect, being particularly critical of purely sensorimotor based accounts of experience (Di Paolo et al., 2017, p.29-32) and the notion of a blind and brittle structural coupling in autopoiesis (Di Paolo, 2005).

Adaptivity (Di Paolo, 2005; Barandiaran and Moreno, 2008), along with the notion of precariousness (Di Paolo et al., 2017; Beer and Di Paolo, 2023), in this context, become key to the attempt to fill the spaces from Varela et al. (1991), by supplementing 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 viability bounds (Di Paolo et al., 2017, p.120-121). This would be accomplished by means of their adaptive capacity, so by actively monitoring, evaluating and regulating their own states and their coupling with the environment with respect to their intrinsic normativity (Di Paolo et al., 2017, p.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 would bring forth a phenomenological world of affective forces (Nave, 2022).

Also from Di Paolo et al. (2017): "Adaptivity is the capability of an autonomous system to respond to tendencies in the trajectories of its states and its relations to the world, so that when these tendencies approach the boundary of its own

viability *the system modulates its coupling with the world* in a way that tends to avert the crossing of this boundary." (emphasis added).

At this point, the problems resurface, however, because even if the overall theoretical 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 (Villalobos and Ward, 2016). In fact, once we reevaluate computational descriptions, we can see that where such modulation occurs we can expect the sorts of reliable regularities suitable for survival, describable as mechanism, and mappable onto input-output relations.

For example, let's consider the idea of adaptivity as a precondition for sense-making — the processes by which external entities or events become meaningful to an agent so that evaluation of the possible consequences of an interaction in terms of viability conditions would endow meaning to it. Albeit alluring, this idea still requires an explicit account of how a basic organism would evaluate its own states and its interactions with its environment.

If evaluations are made by computational processes (by means of coherent structural changes, acting as mechanisms), then the problem is that structural changes are supposedly evaluated in semantic/meaningful grounds, but then, any sort of evaluation would require more structural changes itself, thus implying new evaluations and so on *ad infinitum*. If, on the other hand, we presuppose some phenomenological property, where would this come from outside or beyond cognitive mechanisms? If we were to consider an independent system performing such evaluations, wouldn't that take us to an infinite regress? Otherwise, if monitoring and evaluating can take place by following purely *blind* mechanistic processes, could we still think of a mind? or are we expecting something *more*? If something else is missing, what is it?

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 have not been properly justified (Hutto, 2005; Hutto and Myin, 2020, 2017, chapter 4). 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, solely the consequence of biological evolutionary forces, along with the adaptive ontogeny characteristic of living forms. Hence there is no need for an 'agent-ghost in the machine' guiding behavior by making judgements, 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?"* (Hutto and Myin, 2017, p.111).

To fill the void left by removing the sense-maker from the equation, REC has introduced the concept of *ur-intentionality*, an hypothetical information-based, primitive and goal-oriented tendency to action. Ur-intentionality is thought to have a biologically based normative dimension and to be adaptive during the life of an individual, so even if devoid of (meaningful, contentful) intentionality, it accommodates the notion of directedness by appealing to an organismic sensitivity to environmental features (Hutto and Myin, 2017, chapter 5). In simple words, ur-intentionality can be depicted as a non-contentful correlation of the internal states of an organism an its environment, very close to the idea of a *blind* organic machinery (Gallagher, 2023). Thus, similarly 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. Notwithstanding, the fundamental difference would be given by the posed goal-oriented nature of the actions of the organism as a whole: a fly avoiding an obstacle would not need meanings, but a set of biological norms for global behavior (hence, the importance of the normative dimension) (Hutto and Myin, 2017).

Roughly speaking, the argument is as follows: let's consider the idea that there is no global mind emerging from life itself, so that, at least minimal living systems can be conceived of as incredibly fine tuned evolutionary ensembles of individuals (a sort of colony) that has evolved to act together as a whole, but where there is no collective overall sentient system. Why do they remain together? or better, why/how did they go through such a process of amalgamation? 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. So, while a flock of birds can separate and join again going 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 place to a combined selectivity that progressively overrides/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, that would bootstrap ur-intentionality, because the behavior of the system is now objectively directed towards the distinctions arising from the emerging selectivity. There wouldn't be a need for a global mind whatsoever and there wouldn't be any real intentionality either, not because there's no directedness to-

wards/about an object, but because there's 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 for the origin of cognitive behavior (as of sense-making). Unfortunately, it does not provide us with a concrete explication of the mind, or less even 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, that could become the observer, radical enactivism has left us (even if with a clearer picture of the cognition-to-sentience gap) almost in the same place as before (Rowlands, 2015; Abramova and Villalobos, 2015).

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 urintentionality into meaningful or contentful intentionality, because even if we were to consider a perfectly/fully amalgamated system (hence with absolute global selectivity and directedness, as in autopoiesis) 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.

Back to autopoiesis

From an even more *radical* stance and mainly based on a strict alignment with the strict naturalism defended in the original ideas from autopoiesis (Maturana and Varela, 1973, 1987; Maturana, 2002, 1994), 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 of language (Villalobos, 2013b; Villalobos and Silverman, 2018; Villalobos and Palacios, 2021). The theoretical inclusion of teleology (i.e., purposeful or goal oriented action), underlying notions like adaptivity, normativity or agency imply a kind of system control and regulation, which are based on properties beyond natural phenomena (so beyond their mechanistic biological nature) (Villalobos, 2013b; Villalobos and Ward, 2016) 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 went around the Earth, or of a virus *looking for hosts to invade*. On this thinking, life wouldn't be any different; it may seem to us that living systems act towards their environments, but only because it is natural (for us) to perceive them so, given our preconception of life, besides the intractable complexity of their internal dynamics (Abramova and Villalobos, 2015; Villalobos and Ward, 2016; Villalobos and Palacios, 2021). 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 of Maturana and Varela (1973).

Basically, ATC sees in enactivism a form of wishful thinking about simple organisms, *as if* they were something more than very complex evolutionary machinery. As a matter of fact, as taken from this view, 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 – hence of any kind of intentionality as well (Villalobos, 2013a). Simply put; no mind we can refer to, no active or passive experiencer, no ghost in the machine at all.

Pervasiveness of computation

Perhaps ironically enough, this return to autopoiesis can be seen, within a broader context, as a return to a new form of computationalism.

On the one hand, this is different from the idea of mental representations as envisioned by cognitivist accounts (Newell and Simon, 1976), which required for objects to be *internalized*, hence conceived somehow as mental equivalents of photographs or video recordings. Essentially, cognitive functions were considered to be operating over mirroring symbols of the world *outside*. This, however, would entail that mental representations shared some fundamental (objective) truth across every cognitive system capable of instantiating these symbols (Fodor, 1975; Marr, 1982) in order to correctly internalize them, which seems like a trivial impossibility nowadays. In simple words, the crux of the problem is two-folded, because a system thus conceived not only had to instantiate abstract representations through symbols, but also and fundamentally, their structured predicates had to be logical truths of the real world

Ideas of functional closure as proposed in Maturana (1970); Maturana and Varela (1973, 1987) rule out this paradigm, because the subjective take of any cognitive (and protocognitive) system would make impossible a truthful representation of any kind. As noted by Dell (1985), these systems are of the same kind as those described by Ashby (1956): thermodynamically open (insofar as their processes involve a constant exchange of energy with their surroundings), while informationally closed, in the sense that, because of their self-referential nature, *functionally* speaking, information from their environment is *relational*, therefore fundamentally subjective (Maturana, 1970; Maturana and Varela, 1973; Maturana, 1975, 1987, 2002; Varela, 1979). This is the principle underpinning the idea that autonomous systems have no inputs or outputs, which, however, does entail that their processes cannot be formalized computationally (see Villalobos and Dewhurst (2018) for a more detailed

explanation).

While in autopoietic accounts this was still underdeveloped, leading to a view of absolute closure, by which each system would basically have a fully independent internal perspective (Maturana, 1988; Varela, 1994), enactivism *solves* this conundrum through the concept of enaction, whereby the interpretation a system performs is the consequence of a 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.

It is in this sense, and insofar as any form of consistency of this kind can be formalized as a set consistent regularities, which will therefore require a physical structure that supports it (this is the role of embodied cognition in enactivism), that it can be described and conceptualized as computation (Putnam, 1960; Piccinini and Scarantino, 2011; Villalobos and Dewhurst, 2017; Korbak, 2021).

As explained by Villalobos and Dewhurst (2017); Dewhurst and Villalobos (2017), computation, in this very 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. 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 (Turing, 1950; Copeland, 1993). Along these lines, computation would be the mathematical label or description for consistent mechanisms and processes that appear to be multiply realizable.

This seems bolstered by mounting evidence that computational processes can be realized by chemical reactions in the absence of living or even autonomous systems (Perez-Mercader et al., 2013; Dueñas-Díez and Perez-Mercader, 2019; López-Díaz et al., 2024) and the fact that non-living physical systems are capable of exhibiting behaviors of the kind that we associate with life (Hanczyz and Ikegami, 2010; McGregor and Virgo, 2011), thus hinting that intelligent behavioral patterns precede and are *subsumed* by-, or maybe better, forged in-, living systems (Egbert et al., 2023).

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 (Beer, 1997; Webb, 2001, 2002; Baddeley et al., 2012), reaffirming the point that computation (insofar as mathematical descriptions and algorithmic implementations) seem to capture a fundamental mechanistic dimension of life, also posited by autopoietic theories. All this, within a broader framework oriented to an understanding of the principles of life and cognition in computational or multiple realizable terms (Agüera y Arcas et al., 2003; Sayama, 2008; Razeto-Barry, 2012; Korbak, 2021; Rouleau and Levin, 2023; Goni-Moreno, 2024; López-Díaz et al., 2024), has configured a sort of new computationalist approach, against other positions that consider life to be a special and irreplaceable substrate for the mind (Maturana, 2002; Thompson, 2007; Wiese and Friston, 2021; Seth, 2024).

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.

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. Thus, problems and ideas in principle restricted to information theory and computer science may prove relevant to every sort of intelligent behavior and, hence, to cognition and life after all.

To be clear, and building on the thought experiment proposed by Baltieri et al. (2020); if we were to consider a Watt Governor, would we say that it is performing computations when regulating pressure and heat? As noted in Bermudez (2022), the hypothetical implementation of an algorithmic Watt Governor would be by far worse than just *leaving the device to operate on its design alone*. On the one hand, the conflation between some system's behavioural 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 Turing computable, we do not mean that its physical implementation is necessarily performing computations (perhaps apart from some actual computational device), but that its operation can be 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 description may be far more inefficient!

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 very 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.

In this same vein, we can consider the following example:

a Turing machine, or any automata 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/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 mechanism, insofar as a consistent procedure. For if there were no mechanism to determine the changes of a system (its behavior), there wouldn't 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 needn't 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/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) wouldn't, at least in theory, be suitable for the coherence needed for life and cognition.

Cognition and protocognition

As we have seen so far, there is a fundamental tension at the core of enactivism and much of the rest of cognitive science, concerning the relation between life, mind and cognition. In brief, the cause for this is the bioenactivist purported mental properties attributed to living systems, for intentionality, adaptivity or agency, in theory underpinned by their precarious nature and therefore their need for continuous gathering of energy sources for their dynamic realization (Weber and Varela, 2002; Di Paolo, 2005; Thompson, 2007; Di Paolo et al., 2017; Froese, 2021; Di Paolo et al., 2022).

At least from our point of view, even if this seems sensible, it also seems, as more *radical* positions put it, unnecessary and not properly justifiable if we limit ourselves to naturalist descriptions (i.e., by purely looking at the physical mechanisms present in their processes) (Maturana, 2011; Hutto and Myin, 2017; Villalobos and Palacios, 2021). 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.

In this sense, regarding the hypothetical relation between consciousness and mind, if by mind we understand someone or something that experiences the very fact of its existence somehow, even if minimally; We believe that, given the current available knowledge, it is not only prudent, but also sensible to remain agnostic.

On the other hand, while the theoretical framework of the autopoietic approach to cognition strikes us as the best option to build from, because of its emphasis on the material coherences of cognitive processes(Maturana and Varela, 1973, 1987; Villalobos and Silverman, 2018; Villalobos and Palacios, 2021), by reducing cognition to consistent structural dynamics, it conflates properties that, from our perspective, require conceptual delimitation.

In our view, whatever life is, it is clearly something beyond autonomous dynamics. And it is therefore appropriate to label the specific kind of intelligent behavior organisms exhibit as cognitive behavior, inasmuch as the processes they realize surpass by far simple system-environment coherence and consistent interpretation-response mappings (Lyon, 2006; Allen and Friston, 2018; Lyon et al., 2021; Levin et al., 2021; Wiese and Friston, 2021; Lyon and Cheng, 2023). It does not follow from this, however, that they can be said to have a mind (so sentient and purposeful control and regulation over their own states).

Similarly, to attribute intelligence only to living beings is excessively arbitrary. Moreover, it seems to us that, as many others have hinted or explicitly pointed out (Newell and Simon, 1976; Hanczyz and Ikegami, 2010; McGregor and Virgo, 2011; Dueñas-Díez and Perez-Mercader, 2019; Egbert et al., 2023), 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.

As a matter of fact, a simple proof for this is the vast literature on artificial life and biologically inspired artificial intelligence, that has been able to emulate minimal cognitive-like behavior without the need of an organic *motherboard*, where its application in robotics is especially illustrative (Brooks, 1991; Cliff et al., 1993; Beer, 1997; Quinn, 2001; Baddeley et al., 2012; Egbert and Barandiaran, 2014; Tani, 2017). Few would doubt that the mechanisms underlying the behavior of the systems presented in these works give rise to intelligent behavioral patterns; whether this kind of intelligence is equivalent to that of biological systems (i.e., cognitive), 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's say navigation–, they are poised differently; functionally speaking, the computational problem they specify isn't entirely the same.

Thus, we suggest distinguishing as proto-cognitive prop-

erties the specific properties by which autonomous systems display intelligence, and by autonomy a self-referential property (Varela, 1979, 1984; Varela et al., 1991). Therefore 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 (Hanczyz and Ikegami, 2010; Dueñas-Díez and Perez-Mercader, 2019; Löffler et al., 2019), or by artificial implementations, such as patterns in the Game of Life (Beer, 2014, 2020) and especially Agüera y Arcas et al. (2024) (See Rodriguez-Vergara and Husbands (2025) for further discussion). Put differently, a kind of intelligence that is logically and evolutionarily prior to life, while still constrained by a recursive nature, so 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).

Furthermore, the gap between cognition as the kind of intelligence specific to living beings and the abstract notion of mind, which often appears to be the goal in cognitive science, seems somewhat premature as a research objective. Consequently, instead of moving from what we know of living beings towards more complex cognitive capacities, it may be better to look for intelligent/computable properties already present before, and thus underlying, life.

The field of artificial life, in fact, can be viewed as an amalgam of thought experiments animated by chemical, mathematical, computational and other kinds of models, each exploring different properties for which theoretical understanding is still insufficient, with the aforementioned notion of protocognitive properties as an overarching framework.

Summarizing discussion and the role of Artificial Life

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 that ultimately, the nature of life is something inherently different from such descriptions.

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/consistent sequence of events, these chemical transformations illustrate the link between the notions of (physical) mechanism and (mathematical) computation (Perez-Mercader et al., 2013; Dueñas-Díez and Perez-Mercader, 2019; López-Díaz et al., 2024). From this perspective, it is not strange to think that bases of intelligence are materially *hardcoded* in physico-chemical properties.

Although, unlike living systems, the chemical solutions presented in Dueñas-Díez and Perez-Mercader (2019) quickly dissipate, the point remains that regular, consistent patterns of change in a material universe can be abstracted and formalized as mathematical 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, because protocognitive behavior is essentially an amalgam of multiple elements going through state transitions that follows mechanical causation, producing 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.

This is the same as saying that; (i) a system in the same identical state, when paired with an identical state of its environment, will unavoidably produce an identical transformation (Turing, 1950; Piccinini and Corey, 2021; Perales-Eceiza et al., 2024). And (ii), that if such a system is also operationally closed, the transformations that it will undergo, will make the system to 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 their internal dynamics and their modulation of environmental perturbations. Because 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 (Maturana, 1975; Varela, 1979; Allen and Friston, 2018). This does not entail, however, that everything that a living system is and all it does, can be circumscribed within the computational domain (Searle, 1980; Rubin, 2023; Froese, 2023).

Following the Church-Turing thesis, every operation can be abstracted through the notion of cognitive mechanism (Bermudez, 2022), that is, through a form of 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 (Maturana et al., 1968; Varela, 1997; Dewhurst and Villalobos, 2017).

Thus, insofar as a problem can be well-defined in algorithmic terms, there will be some syntactic machine capa-

ble 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 as well as 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, namely, the problem to be solved will presumably not be the characterization of cognitive/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 (Virgo et al., 2021; Potter and Mitchell, 2022; Biehl and Virgo, 2023; Froese, 2023; Horiguchi et al., 2024).

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 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 (Froese and Taguchi., 2019; Rubin, 2023; Lane, 2024).

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 (Penrose, 1994; Longo, 2009; Louie and Poli, 2011; Froese, 2023). Thus, even if each one of the underlying regularities of cognitive systems were combinations of protocognitive features (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 (Longo, 2009; Perales-Eceiza et al., 2024).

Further, the idea of exploitation of (non-algorithmic) phenomena through cognitive mechanisms provides a hypothetical solution for this contradiction and may provide insights towards a principled account of agency, similar to discussions in previous work on possible roles of chaos or randomness in behavior generation (Shim and Husbands, 2019; Froese, 2023). 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 reexamination of some important assumptions in cognitive science, so it should be approached even more carefully. Indeed, unlike deterministic chaos, undecidability hasn't been demonstrated to exist outside mathematics (i.e. in the physical world) (Perales-Eceiza et al., 2024), in spite of ideas, such as randomness being tacitly assumed 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 as a different, non-reducible kind with its own explanatory level.

From this point of view, the life-mind-computation problem greatly overlaps with the core research motivation of Artificial Life (Aguilar et al., 2014; Boden, 2015; Baltieri et al., 2023; Pigozzi, 2023), insofar as an exploration of the limits of purely mechanistic (autonomous) systems logically prior to life and the investigation of the intelligent properties that can be exhibited by them, such as agency (Moreno and Etxeberria, 2005; Biehl and Virgo, 2023; Froese, 2023; Seifert et al., 2024) or temporality (Varela, 1999; Yamashita and Tani, 2008; Gallagher, 2017; Bogotá and Djebbara, 2023; Rodriguez et al., 2023), to name some. In this context, efforts related to non-conventional computing may prove fruitful in the long run (Ackley and Small, 2014; Stepney, 2012; Broersma et al., 2019).

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 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 endeavours 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 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 is its relation 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. Notwithstanding, through this paper we have tried to shed some light on these concepts, to stress the importance of the relation among them, and to highlight the fact that Artificial Life, as a framework, seems to be particularly well suited for undertaking this challenge.

Acknowledgments

The authors would like to thank the anonymous reviewers for their helpful and insightful commentaries.

References

Abramova, K. and Villalobos, M. (2015). The apparent (ur-)intentionality of living beings and the game of content. *Philosophia*, 43:651–668.

Ackley, D. and Small, T. (2014). Indefinitely scalable computing = artificial life engineering. In *Proceedings of the ALIFE 14: The Fourteenth International Conference on the Synthesis and Simulation of Living Systems. ASME*.

Aguilar, W., Santamaria-Bonfil, G., Froese, T., and Gershenson, C. (2014). The past, present and future of artificial life. *Frontiers in robotics and AI*, 1(8).

Agüera y Arcas, B., Alakuijala, J., Evans, J., Laurie, B., Mordvintsev, A., Niklasson, E., Randazzo, E., and Versari, L. (2024). Computational life: How well-formed, self-replicating programs emerge from simple interaction. *arXiv*, 2406.19108v2.

Agüera y Arcas, B., Fairhall, A. L., and Bialek, W. (2003). Computation in a single neuron: Hodgkin and huxley revisited. *Neural Computation*, 15(8):1715 – 1749.

Allen, M. and Friston, K. (2018). From cognitivism to autopoiesis: towards a computational framework for the embodied mind. *Synthese*, 195:2459–2482.

Ashby, W. (1956). *An introduction to cybernetics*. J. Wiley, New York.

Baddeley, B., Graham, P., Husbands, P., and Philippides, A. (2012). A model of ant route navigation driven by scene familiarity. *PLoS Computational Biology*, 8(1):e1002336.

Baltieri, M., Buckley, C. L., and Bruineberg, J. (2020). Predictions in the eye of the beholder: an active inference account of watt governors. In *ALIFE2020: The 2020 Conference on Artificial Life*.

Baltieri, M., Iizuka, H., Witkowsi, O., Sinapayen, L., and Suzuki, K. (2023). Hybrid life: Integrating biological, artificial, and cognitive systems. *WIREs Cognitive Science*, page e1662.

Barandiaran, X. (2017). Autonomy and enactivism: Towards a theory of sensorimotor autonomous agency. *Topoi*, 36:409–430.

Barandiaran, X., Di Paolo, E., and Rohde, M. (2009). Defining agency: Individuality, normativity, asymmetry, and spatio-temporality in action. *Adaptive Behavior*, 17(5):367–386.

Barandiaran, X. and Moreno, A. (2008). Adaptivity: From metabolism to behavior. *Adaptive Behavior*, 16(5):325–344.

Beer, R. (1995). A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence*, 72:173–215.

Beer, R. (1997). The dynamics of adaptive behavior: A research program. *Artificial Intelligence*, 20:257–289.

Beer, R. (2014). The cognitive domain of glider in the game of life. *Artificial Life*, 20:183–206.

Beer, R. (2020). Bittorio revisited: Structural coupling in the game of life. *Adaptive Behavior*, 28(4):197–212.

Beer, R. and Di Paolo, E. (2023). The theoretical foundations of enaction: Precariousness. *Biosystems*, 223(104823).

Beran, M., Parrish, A., Perdue, B., and Agnes, S. (2014). *Comparative Cognition: Past, Present, and Future*. International Journal of Comparative Psychology (by the authors).

Bermudez, J. L. (2022). *Cognitive Science. An Introduction to the Science of the Mind*. Texas A & M University.

Biehl, M. and Virgo, N. (2023). Interpreting systems as solving pomdps: a step towards a formal understanding of agency. In *Buckley, C.L., et al. Active Inference. IWAI 2022, Communications in Computer and Information Science, vol 1721*. Springer, Cham.

Boden, M. A. (2009). Life and mind. *Minds & Machines*, 19:453–463.

Boden, M. A. (2015). Creativity and alife. *Artificial Life*, 21:354–365.

Bogotá, J. D. and Djebbara, Z. (2023). Time-consciousness in computational phenomenology: a temporal analysis of active inference. *Neuroscience of consciousness*, 2023(1):1–12.

Bowes, S. (2023). *Naturally Minded: Mental Causation, Virtual Machines, and Maps*. Springer Verlag.

Broersma, H., Stepney, S., and Wendin, G. (2019). Computability and complexity of unconventional computing devices. In *Computational Matter. Editors: Susan Stepney, Steen Rasmussen, Martyn Amos*. Springer Cham.

Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence*, 47:139–159.

Bruineberga, J., Dolegab, K., Dewhurst, J., and Baltieri, M. (2022). The emperor's new markov blankets. *Behavioral and Brain Sciences*, 45(e183):1–76.

Buhrmann, T. and Di Paolo, E. (2017). The sense of agency - a phenomenological consequence of enacting sensorimotor schemes. *Phenomenology and the Cognitive Sciences*, 16:207–236.

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3).

Cliff, D., Husbands, P., and Harvey, I. (1993). Explorations in evolutionary robotics. *Adaptive Behavior*, 2(1):73–110.

Copeland, J. (1993). *Artificial Intelligence: A Philosophical Introduction*. Wiley-Blackwell.

Dell, P. (1985). Understanding bateson and maturana: Toward a biological foundation for the social sciences. *Journal of Marital and Family Therapy*, 11(1):1–20.

Dewhurst, J. and Villalobos, M. (2017). The enactive automaton as a computing mechanism. *Thought*, 6:185–192.

Di Paolo, E. (2003). Organismically-inspired robotics: Homeostatic adaptation and natural teleology beyond the closed sensorimotor loop. In *K. Murase & T. Asakura (Eds) Dynamical Systems Approach to Embodiment and Sociality*. Advanced Knowledge International, Adelaide, Australia.

Di Paolo, E. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, 4:429–452.

Di Paolo, E., Burhmann, T., and Barandarian, X. (2017). *Sensorimotor Life: An enactive proposal*. Oxford University Press.

Di Paolo, E. A., Thompson, E., and Beer, R. D. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, 3(2).

Dueñas-Díez, M. and Perez-Mercader, J. (2019). How chemistry computes: Language recognition by non-biochemical chemical automata. from finite automata to turing machines. *iScience*, 19:514–526.

Egbert, M. and Barandiaran, X. (2014). Modeling habits as self-sustaining patterns of sensorimotor behavior. *Frontiers in Human Neuroscience*, 8(590):1–15.

Egbert, M., Hanczyc, M. M., Harvey, I., Virgo, N., Parke, E. C., Froese, T., Sayama8, H., Penn, A. S., and Bartlett, S. (2023). Behaviour and the origin of organisms. *Origins of Life and Evolution of Biospheres*, 53:87–112.

Fodor, J. (1975). *The Language of Thought*. Cambridge MA: Harvard University Press.

Froese, T. (2021). To understand the origin of life we must first understand the role of normativity. *Biosemiotics*, 14:657–663.

Froese, T. (2023). Irruption theory: A novel conceptualization of the enactive account of motivated activity. *Entropy*, 25(5):748.

Froese, T. and Di Paolo, E. (2011). The enactive approach. theoretical sketches from cell to society. *Pragmatics and Cognition*, 19(1):21–36.

Froese, T. and Taguchi., S. (2019). The problem of meaning in ai and robotics: Still with us after all these years. *Philosophies*, 4(14).

Gallagher, S. (2017). The past, present and future of time-consciousness: From husserl to varela and beyond. *Constructivist Foundations*, 13(1):91–97.

Gallagher, S. (2023). *Embodied and Enactive approaches to Cognition*. Cambridge University Press.

Garzon, P. C. and Keijzer, F. (2011). Plants: Adaptive behavior, root-brains, and minimal cognition. *Adaptive Behavior*, 19(3):155–171.

Gershenson, C., Trianni, V., Werfel, J., and Sayama, H. (2020). Self-organization and artificial life. *Artificial Life*, 26:391–408.

Goni-Moreno, A. (2024). Biocomputation: Moving beyond turing with living cellular computers. *Communications of the ACM*, 67(6):70–77.

Hanczyz, M. and Ikegami, T. (2010). Chemical basis for minimal cognition. *Artificial Life*, 16(3):233–243.

Harvey, I., Di Paolo, E., Wood, R., and Quinn, M. (2005). Evolutionary robotics: A new scientific tool for studying cognition. *Artificial Life*, 11:79–98.

Horiguchi, L., Maruyama, N., Shigeto, D., Crosscombe, M., and Ikegami, T. (2024). Quantifying autonomy in ant colonies using non-trivial information closure. In *ALIFE 2024: Proceedings of the 2024 Artificial Life Conference*. MIT Press.

Hutto, D. and Myin, E. (2017). *Evolving Enactivism. Basic Minds Meet Content*. MIT Press.

Hutto, D. D. (2005). Knowing what? radical versus conservative enactivism. *Phenomenology and the Cognitive Sciences*, 4:389–405.

Hutto, D. D. and Myin, E. (2020). Deflating deflationism about mental represen- tation. In *J. Smortchkova, K. Dołrega, & T. Schlicht (eds.), What Are Mental Representations?* Oxford: Oxford University Press.

Jonas, H. (1966). *The phenomenon of life: toward a philosophical biology.* Northwestern University Press, Evanston.

Keijzer, F. (2015). Moving and sensing without input and output: early nervous systems and the origins of the animal sensorimotor organization. *Biol Philos*, 30:311–331.

Kirchhoff, M. and Froese, T. (2017). Where there is life there is mind: In support of a strong life-mind continuity thesis. *Entropy*, 19(4):169.

Korbak, T. (2021). Computational enactivism under the free energy principle. *Synthese*, 198:2743–2763.

Lane, P. A. (2024). Robert rosen's relational biology theory and his emphasis on non-algorithmic approaches to living systems. *Mathematics*, 12:3529.

Levin, M. (2023). Bioelectric networks: the cognitive glue enabling evolutionary scaling from physiology to mind. *Animal Cognition*, 26:1865–1891.

Levin, M., Keijzer, F., Lyon, P., and Arendt, D. (2021). Uncovering cognitive similarities and differences, conservation and innovation. *Phil. Trans. R. Soc. B*, 376:20200458.

Longo, G. (2009). From exact sciences to life phenomena: Following schrödinger and turing on programs, life and causality. *Information and Computation*, 207:545–558.

Louie, A. H. and Poli, R. (2011). The spread of hierarchical cycles. *International Journal of General Systems*, 40(3):237–261.

Lyon, P. (2004). Autopoiesis and knowing: Reflections on maturana's biogenic explanation of cognition. *Cybernetics And Human Knowing*, 11(3):21–46.

Lyon, P. (2006). The biogenic approach to cognition. *Cognitive Processing*, 7:11–29.

Lyon, P. and Cheng, K. (2023). Basal cognition: shifting the center of gravity (again). *Animal Cognition*, 26:1743–1750.

Lyon, P., Keijzer, F., Arendt, D., and Levin, M. (2021). Reframing cognition: getting down to biological basics. *Phil. Trans. R. Soc. B*, 376:20190750.

López-Díaz, A. J., Sayama, H., and Gershenson, C. (2024). The origin and evolution of information handling. *arXiv*, arXiv:2404.04374.

Löffler, R. J. G., Hanczyc, M. M., and Gorecki, J. (2019). A hybrid camphor–camphene wax material for studies on self-propelled motion. *Phys. Chem. Chem. Phys.*, 21:24852–24856.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W. H. Freeman.

Maturana, H. (1970). Biology of cognition. In *Biological Computer Laboratory, BCL Report 9*. University of Illinois, Urbana.

Maturana, H. (1975). The organization of the living: A theory of the living organization. *International Journal of Man-Machine Studies*, 7(3):313–332.

Maturana, H. (1987). Everything said is said by an observer. In *Thompson W. I. (ed.) Gaia: A way of knowing*. Lindisfarne Press, New York.

Maturana, H. (1994). *Preface from Humberto Maturana Romesîn to the second edition. In: De Mâquinas y seres vivos. Autopoiesis: la organizaciôn de lo vivo*, chapter Preface. Editorial Universitaria.

Maturana, H. (2002). Autopoiesis, structural coupling and cognition: A history of these and othe notions in the biology of cognition. *Cybernetics and Human Knowing*, 9(3-4):5–34.

Maturana, H. (2011). Ultrastability... autopoiesis? reflective response to tom froese and john stewart. *Cybernetics and Human Knowing*, 18(1-2):143–152.

Maturana, H., Uribe, G., and Frenk, S. (1968). A biological theory of relativistic colour coding in the primate retina. *Arch Biol Med Exp*, 1:1–30.

Maturana, H. and Varela, F. (1973). *Autopoiesis: the organization of the living. [De maquinas y seres vivos. Autopoiesis: la organizacion de lo vivo]. 7th edition from 1994.* Editorial Universitaria.

Maturana, H. and Varela, F. (1987). *The tree of knowledge: The biological roots of human understanding*. New Science Library/Shambhala Publications.

Maturana, H. R. (1988). Reality: The search for objectivity or the quest for a compelling argument. *The Irish Journal of Psychology*, 9(1):25–82.

McGregor, S. and Virgo, N. (2011). Life and its close relatives. In *Kampis, G., Karsai, I., Szathmáry, E. (eds) Advances in Artificial Life. Darwin Meets von Neumann. ECAL 2009. Lecture Notes in Computer Science(), vol 5778*. Springer, Berlin, Heidelberg.

Merleau-Ponty, M. (2012). *Phenomenology of Perception*. Routledge.

Moreno, A. and Etxeberria, A. (2005). Agency in natural and artificial systems. *Artificial Life*, 11:161–175.

Nave, K. (2022). *Every body's gotta eat: why autonomous systems can't live on prediction-error minimization alone*. PhD thesis, University of Edinburgh.

Newell, A. (1980). Physical symbol systems. *Cognitive Science*, 4(2):135=183.

Newell, A. and Simon, H. A. (1976). Computer science asempirical inquiry: Symbols and search. *Communications of the ACM*, 19(3):113–126.

Penrose, R. (1994). *Shadows of the mind. A search for the missing science of consciousness*. Oxford University Press.

Perales-Eceiza, A., Cubitt, T., Gu, M., and Pérez-García, D. (2024). Undecidability in physics: a review. *arXiv*, 2410.16532v1.

Perez-Mercader, J., Dueñas-Diez, M., and Case, D. (2013). Chemically-operated turing machine. Patent US20140200716A1 United States.

Pezzulo, G., Parr, T., andAndy Clark, P. C., and Friston, K. (2024). Generating meaning: active inference and the scope and limits of passive ai. *Trends in Cognitive Sciences*, 28(2):97–112.

Piccinini, G. and Corey, M. (2021). Computation in physical systems. *The Stanford Encyclopedia of Philosophy (Summer 2021 Edition), Edward N. Zalta (ed.)*.

Piccinini, G. and Scarantino, A. (2011). Information processing, computation, and cognition. *J Biol Phys*, 37(1):1–38.

Pigozzi, F. (2023). Of typewriters and pcs. In *Proceedings of the ALIFE 2023: Ghost in the Machine. MIT Press*.

Potter, H. and Mitchell, K. (2022). Naturalasing agent causation. *Entropy*, 24(472).

Putnam, H. (1960). Minds and machines. In *Dimensions of Mind: A Symposium, S. Hook (ed.)*. New York: Collier.

Quinn, M. (2001). Evolving communication without dedicated communication channels. In *Kelemen, J., Sosík, P. (eds) Advances in Artificial Life. ECAL 2001. Lecture Notes in Computer Science(), vol 2159*. Springer, Berlin, Heidelberg.

Razeto-Barry, P. (2012). Autopoiesis 40 years later. a review and a reformulation. *Orig Life Evol Biosph*, 42:543–567.

Rodriguez, F., Husbands, P., Ghosh, A., and White, B. (2023). Frame by frame? a contrasting research framework for time experience. In *ALIFE 2023: Ghost in the Machine: Proceedings of the 2023 Artificial Life Conference*, page 75. MIT Press.

Rodriguez-Vergara, F. and Husbands, P. (2025). Proto-cognitive bases of agency. *Preprints*, https://doi.org/10.20944/preprints202410.1351.v2.

Rostowski, A. (2022). Freedom: An enactive possibility. *Human Affairs*, 32(4):427–438.

Rouleau, N. and Levin, M. (2023). The multiple realizability of sentience in living systems and beyond. *Cognition and Behavior*, 10(11):1–7.

Rowlands, M. (2015). Hard problems of intentionality. *Philosophia*, 43:741–746.

Rubin, S. (2023). Cartography of the multiple formal systems of molecular autopoiesis: from the biology of cognition and enaction to anticipation and active inference. *BioSystems*, 230(104955).

Sayama, H. (2008). Construction theory, self-replication, and the halting problem. *Complexity*, 13:16–22.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3):417–457.

Seifert, G., Sealander, A., Marzen, S., and Levin, M. (2024). From reinforcement learning to agency: Frameworks for understanding basal cognition. *BioSystems*, 235:105107.

Seth, A. (2024). Conscious artificial intelligence and biological naturalism. *PsyArXiv Preprints*, June 30, 2024.

Seth, A. and Tsakiris, M. (2018). Being a beast machine: the somatic basis of selfhood. *Trends in Cognitive Science*, 22(11):969–981.

Shettleworth, S. J. (2012). *Fundamentals of Comparative Cognition*. Oxford University Press.

Shim, Y. and Husbands, P. (2019). Embodied neuromechanical chaos through homeostatic regulation. *Chaos*, 29(033123):1–17.

Stano, P., Nehaniv, C., Ikegami, T., Damiano, L., and Witkowski, O. (2023). Autopoiesis: Foundations of life, cognition and emergence of self/other. *Biosystems*, 232(105008).

Stepney, S. (2012). Programming unconventional computers: Dynamics, development, self-reference. *Entropy*, 14(0):1939–1952.

Stewart, J. (1996). Cognition = life: Implications for higher-level cognition. *Behavioural Processes*, 35:311–326.

Tani, J. (2017). *Exploring Robotics Minds. Actions, Symbols, and Consciousness as Self-Organizing Dynamic Phenomena*. Oxford University Press.

Thompson, E. (2005). Sensorimotor subjectivity and the enactive approach to experience. *Phenomenology and the Cognitive Sciences*, 4(4):407–427.

Thompson, E. (2007). *Mind in life: biology, phenomenology and the sciences of mind*. Cambridge MA: Harvard University Press.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, New Series*, 59(236).

van Duijn, M., Keijzer, F., and Franken, D. (2006). Principles of minimal cognition casting cognition as sensorimotor coordination. *Adaptive Behavior*, 14(2):157–170.

Varela, F. (1979). *Principles of Biological Autonomy*. North Holland.

Varela, F. (1984). Two principles for self-organization. In *Ulrich, H., Probst, G.J.B. (eds.) Self-Organization and Management of Social Systems*. Springer Series on Synergetics, vol 26. Springer, Berlin, Heidelberg.

Varela, F. (1991). Organism: A meshwork of selfless selves. In *Tauber, A.I. (eds) Organism and the Origins of Self. Boston Studies in the Philosophy of Science, vol 129*. Springer, Dordrech.

Varela, F. (1994). *Preface from Francisco J. Varela Garcia to the second edition. In: De Mâquinas y seres vivos. Autopoiesis: la organizaciôn de lo vivo*, chapter Preface. Editorial Universitaria.

Varela, F. (1997). Patterns of life: Intertwining identity and cognition. *Brain cognition*, 34(1):72–87.

Varela, F. (1999). Present-time consciousness. *Journal of consciousness studies*, 6(2-3):111–140.

Varela, F., Thompson, E., and Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. The MIT Press.

Villalobos, M. (2013a). Autopoiesis, life, mind and cognition: Bases for a proper naturalistic continuity. *Biosemiotics*, 6:379–391.

Villalobos, M. (2013b). Enactive cognitive science: revisionism or revolution? *Adaptive Behavior*, 21(3):159–167.

Villalobos, M. and Dewhurst, J. (2017). Why post-cognitivism does not (necessarily) entail anti- computationalism. *Adaptive Behavior*, 25(3):117–128.

Villalobos, M. and Dewhurst, J. (2018). Enactive autonomy in computational systems. *Synthese*, 195:1891–1908.

Villalobos, M. and Palacios, S. (2021). Autopoietic theory, enactivism, and their incommensurable marks of the cognitive. *Synthese*, 198(supol. 1):571–587.

Villalobos, M. and Silverman, D. (2018). Extended functionalism, radical enactivism and the autopoietic theory of cognition: prospects for a full revolution in cognitive science. *Phenomenology and the Cognitive Sciences*, 17:719–739.

Villalobos, M. and Ward, D. (2016). lived experience and cognitive science. reappraising enactivism's jonasian turn. *Constructivist Foundations*, 11(2):802–831.

Virgo, N., Biehl, M., and McGregor, S. (2021). Interpreting dynamical systems as bayesian reasoners. In *Kamp, M., et al. Machine Learning and Principles and Practice of Knowledge Discovery in Databases. ECML PKDD 2021. Communications in Computer and Information Science, vol 1524*. Springer, Cham.

Ward, M., Silverman, D., and Villalobos, M. (2017). Introduction: The varieties of enactivism. *Topoi*, 36:365–375.

Wasserman, E. A. (1993). Comparative cognition: Beginning the second century of the study of animal intelligence. *Psychological Bulletin*, 113(2):211–228.

Webb, B. (2001). Can robots make good models of biological behaviour? *Behavioral and Brain Sciences*, 24:1033–1050.

Webb, B. (2002). Robots in invertebrate neuroscience. *Nature*, 417(16 May 2002):359–363.

Weber, A. and Varela, F. (2002). Life after kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences*, 1:97–125.

Wiese, W. and Friston, K. (2021). Examining the continuity between life and mind: Is there a continuity between autopoietic intentionality and representationality? *Philosophies*, 6(18).

Yamashita, Y. and Tani, J. (2008). Emergence of functional hierarchy in a multiple timescale neural network model: A humanoid robot experiment. *PLoS Computational Biology*, 4(11):e1000220.