Why Artificial Life and Consciousness Are Impossible in Principle

Path Dependence, Operative Closure, and the Epistemic-Ontological Structure of the Living

Preliminary Remark

The following essay develops an argument that is, in its very design, at once epistemological and ontological. It does not claim that artificial life or artificial consciousness would be hard to produce, practically improbable, or blocked by the present state of technology. It claims a principled impossibility that follows from the structure of the living itself, and it does so from the convergence of several independent arguments: from the nature of scientific abstraction, from the complexity of the evolutionary space of possibilities, from the active path-production of living systems, from the operative closure of biological autocatalysis, and finally from the ontological thesis that ontology is the epistemics of the ontic. These arguments are not merely additive; they support one another and converge in a central aporia that structurally affects every project of production.

1. Abstraction as the Constitutive Form of Cognition

All scientific knowledge is abstraction. This is not a deficiency that could be remedied by improved methods; it is the constitutive form of cognition itself. When scientists describe a phenomenon, they do not perceive the thing in its full physical concreteness and write it down; they produce a map that foregrounds certain features of the landscape and leaves others behind. This holds for every scientific discipline without exception: physics describes not bodies but functional relations between measurable quantities; chemistry describes not molecules in their singular physical realization but reaction types, energy profiles, bonding orders; biology describes not organisms as concrete historical individuals but processes, regulatory mechanisms, selection pressures, developmental logics.

The philosophical tradition has thematized this basic structure of cognition under various descriptions: as the distinction between the thing-in-itself and appearance, as the difference between the ontic and its epistemic accessibility, as the tension between the concrete and the conceptual. What is decisive is that abstraction represents not an accidental incompleteness but a structural necessity: cognition is generalization, and generalization leaves the particular behind.

This finding is unproblematic across large areas of science. Physics can calculate bridges without seeing the landscape of gravitation in its microscopic complexity, because what is causally relevant hangs on the function, not on the singular physical realization. The abstractum "F = ma" carries everything needed for prediction. The bridge-builder need not know which path the physical field configurations take at the Planck scale; the invariants he uses are fully contained in their abstracted relations. This is possible because the relevant causal structure has no history that would be lost in the process of abstraction: gravitation operates today by the same laws as four billion years ago, and the laws are not constituted by the history of their instantiations.

It is precisely this feature that is structurally different in the case of life and consciousness. The thesis developed in what follows is: the relevant causal structure of living systems is inseparable from its history, from the concrete path that evolution took through the space of possibilities. And this path is contained in no abstraction, not because the abstraction is incomplete, but because abstraction as such leaves the path behind.

2. Functions as Abstracta and the Infinite Space of Possibilities

When biologists, cognitive scientists, and neuroscientists describe life and consciousness, they describe functions. Functions are abstracta in a precise logical sense: they specify a mapping relation between input and output states and leave open by which concrete mechanism this relation is realized. "The heart pumps blood" is a functional description compatible with a very large number of physically different realizations. "Neural excitation transmission," "adaptive response to environmental stimuli," "self-preserving behavior," "phenomenal experience as integration of information" are functional descriptions, each of which opens a space of possible realizations.

This space of possibilities has a decisive mathematical property: it is not finite. Already with a few interacting elements, combinatorics generates a number of possible configurations that must, for any practical purpose, be treated as infinite. Stuart Kauffman has shown this for autocatalytic sets: already with a relatively small number of different molecule types, the number of possible reaction networks that realize the same autocatalytic closure exceeds any realistic horizon of exploration. Ilya Prigogine and Isabelle Stengers have shown something similar for dissipative structures: the bifurcation points at which a far-from-equilibrium system chooses among possible states of order generate branching trees whose complexity grows not linearly but superexponentially with system size.

It is necessary at this point to determine precisely what these references do and do not accomplish for the present argument. Kauffman supplies the combinatorial argument: the sheer size of the space of possible autocatalytic networks. Prigogine and Stengers supply the bifurcation argument: the branching structure through which a non-equilibrium system, at critical points, realizes one order among several possible ones. Both arguments establish the size and the branchedness of the space of possibilities, and both are necessary for the thesis pursued here. But neither of them, on its own, already carries the historicity thesis that is ultimately at stake. A large space of possibilities is still a space, that is, a pre-given structure through which a path leads; and a bifurcation tree, however branched, is a describable-in-advance topology of possible paths. In both versions the space of possibilities remains the logically prior, the path the logically posterior, which selects one realization out of the space.

The thesis that carries the present argument is stronger and goes beyond both Kauffman and Prigogine: that the space of possibilities of living systems is not given in advance, through which a path would then run, but is continually co-produced by the activity of the system, so that path and space cannot be separated as the logically prior and the logically posterior. Kauffman's combinatorics and Prigogine's bifurcations describe the size and branchedness of a space they treat as pre-given; the argument developed here disputes precisely this pre-givenness. In this respect these references are taken up not as a foundation on which the thesis would rest, but as a description of an intermediate stage of the argument that has not yet reached the actual thesis, the co-production of the space of possibilities by the active system. They show how large the space would be if it were a space; the actual point is that it is none in the pre-given sense.

Evolution has realized exactly one path out of this infinite space of possibilities. This sentence carries more than it seems to at first glance. It means not only that the evolutionary path is one among many possible ones; it means that the number of possible paths is so large that no search procedure that does not itself possess the same historical depth and the same causally recursive structure as evolution itself could reproduce it. The impossibility is not meant practically, in the sense of "too costly given available resources"; it is a matter of the relations between the space of possibilities and any finite procedure of exploration, and these relations are structural.

What do we know of the path that evolution took? We know the statistical invariants it has left behind: the regularities that recur across many instances of biological systems, metabolic scaling laws, regulatory logics, information-processing patterns. These invariants are real. But they are not the path; they are what many possible paths have in common, that is, precisely what does not determine the concrete path. To infer the path from the invariants is structurally impossible, because many different paths could have generated the same invariants, and because the invariants arise through abstraction from the path, that is, through a process that systematically erases the particularity of the path.

3. The Collapse of Epistemology and Ontology

At this point there is an obvious objection: the barrier described is epistemic, not ontological. The evolutionary path is real; it took place; it lies in the past. Our ignorance is a cognitive deficit, not a structural feature of the matter. And where an epistemic barrier is present, remedy could in theory be found through better instruments, deeper analysis, more comprehensive data.

This objection presupposes a dichotomy that must itself be questioned: the dichotomy between a world that is as it is, independently of its being known, and a knowing subject that maps this world with more or less success. This dichotomy is untenable in its naive form as soon as one understands ontology as the epistemics of the ontic. Ontic determinations are not given pre-cognitively, as though they had a mode of givenness independent of all cognition; they are the result of cognition-forming processes that are themselves historical, situated, and constituted by their own prehistory. There is no ontic datum that has not passed through a knowing system, through a process of constitution, selection, abstraction.

For the present context this means: the epistemic barrier with respect to the evolutionary path is not a barrier of our knowledge concerning a state of affairs accessible in itself. It is a barrier belonging to the state of affairs itself, because the state of affairs is present only in the form in which it can be constituted by knowing systems, and this form is abstraction, the map, not the landscape. The path in its concrete physical singularity lies nowhere in a form that would make it accessible from the outside. It is not hidden like a treasure one could in principle find if one only dug deep enough. Its inaccessibility is constitutive of its status as the path of a self-organizing, operatively closed, historically constituted system.

The first and the second argument, the epistemic and the ontological, thus collapse into one another. The impossibility of knowledge about the path is not an accidental property of our cognitive equipment; it is a property of the matter, because the matter is so constituted that it is accessible only as a map, and because the path is precisely what the map leaves behind.

4. Evolution as an Active System: Path-Production Instead of Path-Traversal

A further objection to the impossibility thesis comes from the theory of evolution itself. The standard account of evolutionary dynamics describes random variation and non-random selection. If that is correct, evolution works through a blind search procedure with feedback; and if a blind procedure without knowledge of the space of possibilities can produce consciousness, then another blind procedure, such as a genetic algorithm with suitable fitness criteria, could in principle do the same.

This objection rests on a misreading of the relation between evolution and its space of possibilities. Evolution does not traverse a pre-given, stable space of possibilities with random steps. Through its own activity it co-produces the space of possibilities through which it moves. Living systems alter their selective environment through their behavior: they create new ecological niches, shift resource availabilities, generate new predator-prey relations, modify the chemical and physical milieu. This process, known as niche construction, systematically elaborated by Richard Lewontin and playing a central role in the theory of developmental plasticity, means that the topology of the space of possibilities is not stable with respect to the steps that lead through it. Every step alters the landscape in which the next step takes place.

What is decisive here is the following: evolution is not blind in the sense in which a search algorithm over a pre-given search space is blind. It is active in the sense that its activity co-produces the search landscape itself. Herein lies the systematic difference from any technical search procedure, which always operates on a pre-given search space that must be defined before the search begins. The search space of a genetic algorithm is fixed by its constructor; the "search space" of evolution is continually reconstituted by evolution itself and was not definable before the first step, because it arises only through the steps.

The trial-and-error procedure of evolution is therefore not comparable to the trial-and-error procedure of an engineer optimizing a technical system. It is a procedure that continually reconstitutes the very field of meaning of "trial" and "error," because what counts as error is defined by the altered selective environment that the previous trial helped to produce. The procedure is not transparent from our perspective, because we cannot survey from the outside the space of possibilities through which it moves: it is not stable enough to be mapped.

5. Operative Closure: The Constitutive Difference Between Chemical and Biological Autocatalysis

The sharpest structural difference between living and non-living systems lies in operative closure. To make this concept productive, it must be demarcated precisely against an obvious precursor.

A chemical autocatalytic system, such as occurs in Belousov-Zhabotinsky reactions or in simple autocatalytic cycles, is self-amplifying in the sense that reaction products function as catalysts for the same or cooperating reactions. It thereby realizes an internal feedback structure that distinguishes its dynamics from the linear kinetics of simple reactions. But the boundary conditions of this system, the concentrations of the starting materials, the temperature, the pH value, the spatial extent of the reaction vessel, are set and controlled from the outside. The system is circular in itself, but its boundary is not itself a product of its dynamics. Someone filled the reaction vessel; someone keeps the temperature constant; someone defines what counts as the system boundary. The self-referentiality of the system is a limited one, made possible externally.

A biological autocatalytic system is operatively closed in a more fundamental sense: it produces not only its own components in a circular network of reactions, but also the boundary that constitutes it as a system, and the processes by which it organizes its own production. The cell membrane is not an external container; it is a product of the cellular processes and at the same time the condition under which these processes can take place. The system and its boundary co-produce one another in a process that has no external instance setting the boundary conditions.

The concept of operative closure was fundamentally elaborated by Humberto Maturana and Francisco Varela in their theory of autopoietic systems, and the present argument takes from there the basic insight that a living system circularly brings forth its own organization and its own boundary. At a decisive point, however, the argument developed here goes beyond the classical conception of autopoiesis, in a way that represents not merely a supplement but a shift of emphasis. In Maturana and Varela, operative closure is conceived primarily organizationally and, in this sense, ahistorically: autopoiesis is the invariant organization that distinguishes a system as a living system, and this organization remains equal to itself as long as the system lives. The historicity that is load-bearing for the present argument plays no constitutive role in this version; the autopoietic organization is precisely what remains the same across the system's history.

The argument developed here shifts the emphasis from the invariant organization to the historically constituted path. What is causally relevant is not the organizational form as such, but the fact that this organizational form is, at every moment, the result of all its previous states, and that the boundary conditions under which it operates are internal and historical. Operative closure is thus not an ahistorical invariant that distinguishes a system as living, but the mechanism by which a system enfolds its own history into its present organization, without that history being any longer legible as such in the present organization. This shift is not cosmetic: it is what connects operative closure with path dependence and non-factorizability, turning an organizational concept into a historical-causal one.

At the same time, the operatively closed system is not thermodynamically closed. It is an open system that exchanges matter and energy with its environment; without this exchange it would fall prey to thermodynamic dissipation. Operative closure concerns not the exchange of matter but the organizational form: through its internal structure, the system determines how it processes external thermodynamic flows, which perturbations it takes up as operatively relevant and which not, which disturbances it treats as disturbances and which it absorbs as irrelevant noise. This self-determination of relevance is not determined by the external flows themselves; it is a function of the internal organization, which is in turn the result of all previous states.

The living system thus operates in a tension that has no stable middle position: between the intrinsic dynamics defined by the internal organization and the thermodynamic flows acting from outside. It is forced into this tension, because it would cease to exist if it acted either exclusively internally (thermodynamic death through entropy increase) or were determined exclusively externally (dissolution of the organizational form through overwhelming of the internal regulatory capacity). Self-organization is not an optimum the system strives for; it is a necessity it cannot evade as long as it lives. This "being forced" into self-organization is constitutive of the status of a living system.

The structural difference between chemical and biological autocatalysis is thereby clear: in the chemical case, boundary conditions can be controlled because they are not part of the system dynamics. In the biological case, the boundary itself is a product of the internal organization, and this organization in turn changes in the very way in which it processes the external flows. There is no stable interface that could be isolated and reproduced, because the interface itself is continually produced anew by exactly the process it bounds.

6. Statistical Invariants as Constitutively Retrospective Surfaces

What is scientifically accessible under these conditions? The answer must be formulated precisely, because it is not epistemically trivial.

What is accessible are statistical invariants. Living systems display regularities that recur across many instances and across evolutionary kinship groups: metabolic scaling laws of the following form:

$$P \sim M^{\frac{3}{4}}$$

Information-processing capacities under uncertainty, self-preserving behavior, reproduction patterns, neural regulatory structures. These regularities are epistemically real; they are not mere artifacts of description but describe genuine properties of living systems.

But these properties have an epistemological status that must be carefully determined. Statistical invariants are constitutively retrospective. They arise through comparison across a class of events or systems; they describe what recurs within the set of realized paths, that is, what many paths have in common. They are the projection of a set of paths onto a plane on which path difference has been erased. Their epistemic value lies precisely in this erasure, because the erasure enables comparison, classification, and weak prediction. And their epistemic limit lies likewise in this erasure, because the path itself is precisely what is left behind by abstraction.

The decisive point is this: more invariants, finer invariants, invariants at a lower level of description do not come closer to the path. They describe more surface, but the path does not lie deeper in the structure such that it would become visible under sufficiently fine-resolving analysis. It lies in the history, and history is in principle not contained in the state, because the state is the result of the history, not the history itself. To interrogate a historically constituted state with regard to its history of emergence is to ask about what the state itself no longer contains.

This distinguishes biological invariants systematically from physical conservation laws. A conservation law such as the constancy of total energy describes a property visible in every state of the system, at every time, independently of the history. It is not retrospective in the sense that one would first have to collect many states to see it; it lies in the structure of the physical space of possibilities itself. Biological invariants, by contrast, arise through selection over historical paths and are therefore constitutively retrospective: they describe what evolution has read out as stable, and are thus essentially protocols of a history, not descriptions of an ahistorical structure.

This difference has a direct implication for any research program that pursues artificial life or consciousness through the identification and reimplementation of statistical regularities. Such a program reconstructs the surface, not what produces the surface. It implements a further instance of a map, not a new landscape. Whether the phenomenon arises in the process depends on whether the phenomenon hangs on the surface or on what produces it.

7. The Dynamic Attractor as an Agent: Self-Co-Production of the State Space

Physical and many chemical systems tend toward static attractors in the sense of dynamical systems theory: fixed points, limit cycles, strange attractors in chaotic systems. These attractors lie in the phase space of the system; they can be described from the outside; the system tends toward them without the tendency altering the attractor itself. One can know the attractor of a system without intervening in the system, and knowledge of the attractor enables predictions about long-term behavior, at least in the statistical sense.

Living systems have dynamic attractors in a fundamentally different sense. The internal organization, that is, what governs the behavior of the system at every moment, is not stable in the sense of a fixed region of phase space; it changes with every step of the system and with every interaction with the environment. The system is not something that tends toward an attractor lying outside the system; the system is itself the attractor, and this attractor continually generates itself anew. The system acts, instead of being pulled. It co-produces the space of its possible states by acting within that space.

This self-co-production of the state space has a direct consequence for predictability. When one wishes to say of a human subject that one does not know what it will do or think next, this is not merely an expression of limited information that could be remedied by more data. It is the expression of a principled structure: what the subject will do next is not determined within a pre-given state space, but is co-determined by the acting itself, which co-produces the state space for the next step.

A complete prediction of the next step would, if it were accessible to the subject, itself be an event in the subject's environment and thus an intervention in the state space that the prediction was meant to describe. The subject would react to the prediction: it would confirm, reject, undercut, or ignore it. Each of these reactions would be a further action of the operatively closed system, which processes the external prediction through its internal organization. The prediction is not a neutral describing of an external state of affairs; it is an event that intervenes in the described state of affairs, because the described state of affairs is a system that actively constitutes its own environment.

Statistics remains the only form of approximation. One can give probability distributions over past behavior; one can, on the basis of accumulated observations, formulate expectations about future behavior. But these distributions describe no causal structure that determines the next event in such a way that one could specify it in advance. They describe what was frequent in the past and project this as a weak expectation into the future. The subject can fulfill this expectation, disappoint it, or undercut it through actions lying outside the statistical model, and that not as a random exception but as an expression of its capacity to generate new situations that the statistical model does not contain, because the model is based on the past of the system, not on the ongoing self-co-production of its space of possibilities.

8. The Fundamental Aporia of Construction: Building Presupposes a Map, Life Produces the Landscape

All the foregoing considerations converge in a central aporia that concerns the entire project of producing artificial life or consciousness.

Building presupposes a map. This is no metaphor and no weak claim; it is a logical necessity. To produce a system, one must specify its target state, either fully or sufficiently, and one must give steps that lead from a current state to this target state. Both require descriptions, that is, abstractions, that is, maps. One cannot build a system without having a description of that system, even if the description is implicit, embodied, or present as practical operational knowledge. The engineer who constructs a machine always has a design, a specification, a conception of the goal, even if that conception is vague and changes in the course of construction. Construction is the implementation of an abstractum.

Living systems produce the landscape in which maps first become possible after the fact. This is the asymmetric core structure of the problem. The abstraction that makes it possible to describe a living system is an abstraction from a system that itself never lay before us as an abstraction. It came into being through operative closure, through active path-production, through the inevitability of self-organization under thermodynamic flows, and through the property of continually co-producing its own space of possibilities. None of these processes is itself the implementation of a description. None has a target state toward which it tends that would have lain before the beginning of the process in a form that would have made its construction possible. The "goal" of evolution is not a state that can be specified by a design; it is life itself, in the form in which it is constituted, at every moment of its history, through self-organization.

The aporia is as follows: every reimplementation of a functional description is the production of a new map, and a map is, in its form, factorized. It decomposes the phenomenon into isolable components, functions, relations, parameters, which can be specified independently of one another and reassembled in the construction. This factorizability is not a contingent property of descriptions; it is the condition of their constructibility, for to construct means to assemble a whole from specified parts according to a specified order. Without decomposition into parts determined prior to assembly, there is no construction.

The living system, however, is not factorizable. Its causally relevant structure lies not in isolable components that could be determined independently and reassembled, but in the way it brings forth its own boundary anew at every moment out of the tension between intrinsic dynamics and external flows. This way is not a part among other parts; it is the manner in which the parts first become parts of this system at all. If one decomposes the system into factorizable components, one loses precisely what was to be explained, because non-factorizability is not a technical obstacle of analysis but the organizational form of the system itself. What is lost through factorization is not a detail that could be regained by finer decomposition, but the connectedness whose non-decomposability constitutes the phenomenon.

With this the argument closes, without needing any additional assumption about the irreducibility of phenomenal experience to function. One need not already have believed that there is a phenomenal something, separable from function, that the reimplementation would miss. The impossibility follows one step earlier: construction presupposes a factorized description, but what is causally relevant in the living is precisely its non-factorizability, and a factorized description of a non-factorizable state of affairs leaves the relevant not as a detail but as a whole behind. Building presupposes a map; the map factorizes; the living is what disappears through factorization. This is not a thesis about a hidden extra beyond function, but a thesis about the logical form of construction and the organizational form of life, and about their incompatibility.

9. Concluding Remark: The Structural Limit as an Epistemic Consequence

The result of this argument is not skeptical in a paralyzing sense. Scientific descriptions of living systems are possible, productive, and necessary; statistical invariants have explanatory value; functional models enable interventions and therapies; evolutionary models illuminate the historical logic of biological organizations. None of this is called into question by the thesis developed here.

What is called into question is a methodological basic assumption implicitly at work in many projects of AI research, synthetic biology, and cognitive science: that the complete description of a phenomenon is in principle sufficient to produce it; that the gulf between map and landscape can be bridged by successive refinement of the map; that statistical invariants, at sufficient resolution and depth, fully carry the relevant causal structure.

This assumption is justified when what is causally relevant lies entirely in the abstracted relations. For phenomena whose relevant causal structure has no history, that is, for physical systems operating by ahistorical laws, abstraction grasps the essential. For phenomena whose causally relevant feature is operative closure, path dependence, the active co-production of the space of possibilities, and the historical constitution of the organizational form, the assumption is structurally mistaken.

The insight that follows from this analysis is not one that paralyzes science; it is one that enables a more precise self-knowledge. Maps are indispensable. But it is an epistemic virtue to know what maps leave behind: not only details, but the path, and with it the phenomenon that this path, and only this path, has constituted.

Wolfgang Stegemann, 2026