Synthese (2024) 204:75
https://doi.org/10.1007/s11229-024-04705-w

ORIGINAL RESEARCH

Origins of biological teleology: how constraints represent ends

Miguel García-Valdecasas[1,2] · **Terrence W. Deacon**[3]

Received: 3 September 2022 / Accepted: 10 July 2024
© The Author(s) 2024

Abstract

To naturalize the concept of teleological causality in biology it is not enough to avoid assuming backward causation or positing the existence of an inscrutable teleological essence like the *élan vital.* We must also specify how the causality of organisms is distinct from the causality of designed artifacts like thermostats or asymmetrically oriented processes like the ubiquitous increase of entropy. Historically, the concept of teleological causality in biology has been based on an analogy to the familiar experience of purposeful action. This is experienced by us as a disposition to achieve a general type of end that is represented in advance, and which regulates the selection of efficient means to achieve it. Inspired by this analogy, to bridge the gap between biology and human agency we describe a simple molecular process called *autogenesis* that shows how two linked complementary self-organizing processes can give rise to higher-order relations that resemble purposeful dispositions, though expressed in terms of constraints on molecular processes. Because the autogenic model is described in sufficient detail to be empirically realizable, it provides a proof of principle demonstrating a simple form of teleological causality.

Keywords Teleology · Representation · Constraint · Self · Causality · Organization · Normativity · Autogenesis · Teleodynamics

✉ Miguel García-Valdecasas
garciaval@unav.es

Terrence W. Deacon
deacon@berkeley.edu

1 University of Navarra, Institute for Culture and Society, Campus Universitario, Pamplona 31009, Spain

2 University of Navarra, Facultad de Filosofía y Letras, Campus Universitario, Pamplona 31009, Spain

3 Anthropology Department & Cognitive and Brain Sciences Institute, University of California, Berkeley, Anthropology and Art Practice Building, Berkeley, CA 94720-3710, USA

Published online: 12 August 2024

> … no one has yet succeeded in offering an account of how function, purpose, or agency might emerge from the dynamics of effectively homogeneous systems of simple elements, no matter how complex those dynamics might be.
> Evelyn Fox Keller (2009, p. 27).

1 Introduction

Throughout most of the last two centuries, the natural sciences have systematically shunned teleological explanations. This exclusionary stance is neither surprising nor problematic. Teleological explanations generally lack an account of their mechanism of action and have the superficial appearance of implying backwards causality. Despite these difficulties with the concept and centuries of efforts to replace teleological accounts of biological processes with purely mechanistic accounts, it has proven impossible to purge teleological notions of organism form and activities from biological explanation. Many attempts to reduce teleology to more basic forms of explanations or eliminate it altogether have failed to convince many of the wisdom of this strategy (Jonas, 1966; Woodfield, 1976; Jacobs, 1986; Bedau, 1991, 1992; Hacker, 2007; Thompson, 2007; Deacon, 2012; Moreno & Mossio, 2015; Walsh, 2015; Nguyen, 2021).

Historically, the concept of teleological causality in biology has been based on an analogy to the familiar experience of purposeful action. Human agents experience purposeful action as a disposition to achieve a general type of end that is represented in advance, and which regulates the selection of efficient means to achieve it. The mental representation does not specify in all its details how this end will be physically realized, only that the form that is thereby produced will approximately resemble the content of that prior representation. For most theories of action, the challenge has always been to explain the physical nature of this pre-specification of the end and how it is able to organize work to increase the probability that its form is realized. In this respect, a naturalized theory of teleological causality assumes a naturalized theory of representation.

Purposive action in humans almost certainly has its precursors in the prior non-mental biological mode of agency that is exemplified by an organism's capacity to respond and adapt to its local environment. In the light of evolution, mental agency can be understood as an evolved elaboration of non-mentalistic cellular-molecular processes. But that means that our most familiar version of teleological causality is a special case, not the general form of end-directed behavior. So as not to put the cart before the horse, we endeavor only to explore this most minimal conception of biological teleology.

We recognize the philosophical difficulty that this poses. The concepts of teleology and representation are among the oldest and most contested topics in the history of philosophy. Given this difficulty, rather than resolving the thorny problem of how these concepts relate, we focus on a narrow biological correlate of these problems: how teleological causality can emerge from the dynamics of molecular interactions in the critical transition from chemistry to life.

Before we proceed, let us introduce two caveats. First, we use the term "representation" in a more generic sense than it is standardly found in philosophy and psychology to conform to standard biological use. For example, at least since Francis Crick proposed his so-called "central dogma" of molecular biology, in molecular biology the sequence of nucleotides in a DNA molecule is commonly described as "representing" a sequence of amino acids comprising a protein. It is uncontroversial that the constraints on molecular structures and interactions of DNA provide a representation of the target "design" for ongoing organism functions as well as that of future progeny. This usage assumes only a functional correspondence whereby the nucleotide sequence serves as a recording of a pattern that the organism uses to preserve and transmit constraints on protein structure, and carries no mentalistic connotation.

Second, though teleological explanations in biology are rife, it is often believed that evolution by natural selection has made teleological causality redundant. E.g., in his preface to a new edition of Darwin's work on orchids Michael Ghiselin argues that Darwin's (1959) theory succeeded in "getting rid of teleology and replacing it with a new way of thinking about adaptation." (Darwin, 1862/1877 [1984], p. 409). Similarly, theories of selected effects, extensively developed by Neander, Millikan, Griffith, Godfrey-Smith, and others, assume that the apparent end-directedness of biological functions is not an intrinsic feature of organisms, but rather an observer's projection of whatever properties past replicas of that trait contributed to its current presence in the population. We disagree. In this essay, we do not address the extensive literature arguing that a natural selection-based etiological analysis is enough to understand teleology[1]. Rather, we begin with the assumption that end-directed causality is an undeniable feature of living organisms, and endeavor to demonstrate how it is realized in systems that are much simpler and less prevalent than organisms.

This paper is divided into six sections.

The second section provides a brief overview of previous efforts to characterize the nature of basic organism teleology and outlines six alternative ways of characterizing teleological causality. Among the listed alternatives, autogenesis, the model that we will put forward, reflects a *constitutive* versus a descriptive characterization of teleological causality, a perspective on the locus of teleological agency that is neither internalist nor externalist, a *targeted* versus terminal conception of causal directionality, an intrinsically *normative* versus nonnormative relationship between the source of teleology and its end, and a *general* versus particular characterization of ends. We argue that the use of these categories can help to clarify the underlying assumptions that underpin any teleological theory, and in each case, we identify the alternative that is most consistent with the approach we propose. We do not claim that these are the only relevant alternatives under which teleology can be categorized, or that these are exhaustive. However, we do argue that any teleological theory must in some way or other address these criteria.

This section also distinguishes between asymmetric causal processes that spontaneously develop toward what we call "terminal" states, i.e. those beyond which change is blocked, and causal processes that develop toward what we call "target" states, i.e. those that are reached because of their value to some benefitting entity,

[1] See García-Valdecasas & Deacon, 2024 for a critical engagement with this literature.

irrespective of any spontaneous tendency to reach this state. We recognize only the latter, which involve work to counter terminal tendencies, as teleological and explain why.

The third section introduces the concept of *constraint* and explains its relevance to thermodynamic work. Constraints provide a way for one general form to bring about another general form because of the way work is channeled.

The fourth section discusses the relationship between generals and particulars inspired by C. S. Peirce's characterization. Peirce described final causality as a *general* mode of determination rather than one producing particular physical consequences. We will show, however, that Peirce's characterization is insufficient to ground this idea on the notion of constraint.

The fifth section describes an empirically testable model that embodies the basic criteria for teleological causality specified above. Termed *autogenesis* by Deacon (2012: ch. 10; 2021), it describes two complementary self-organizing processes that provide the essential boundary conditions that each requires in order to persist. In so doing, autogenesis demonstrates the critical role played by a multiply realizable constraint —which we call a *hologenic* constraint— that establishes the individuated unity of the source of this causal disposition. In this way, this constraint creates a differentiated individual. Formally, we argue that individuating the source of end-directed work provides a specific beneficiary of this disposition.

The sixth section describes how this multiply realizable constraint on the reciprocity between lower-order physical constraint-generating processes can provide a form of biological representation and a locus for producing its own future instantiation in new physical substrates.

The seventh section compares autogenesis to the three most popular paradigms dealing with the nonlife/life distinction, which we loosely categorize as replication-based, self-organization-based, and autonomy-based theories, respectively. We present the reasons why autogenesis is a preferable candidate to account for the emergence of end-directed organizations.

The conclusion redescribes the boundaries between basic biological teleology and mental teleological causality, characterizing how the latter has evolved from the former and ultimately depends on it.

2 Basic desiderata for a theory of biological causality

Among the so-called "teleonaturalist" teleological theories (Allen and Neal, 2020), that is, theories for which the truth conditions of biological claims must be sought in non-mental facts about organisms, or living systems in general, we might map out at least five existing conceptual dichotomies concerning teleology. These dichotomies are not meant to be exhaustive, e.g., they do not include eliminative theories, nor do they wade into the contemporary function debate—which tends to focus on the function of biological traits rather than on teleological causality per se. Instead, we take a perspective that specifically focuses on categories deemed relevant to explain teleology as a mode of physical causation and leave many other considerations aside.

2.1 Internalist vs. externalist accounts

According to Aristotle, all nature is pervaded with ends—"nature makes nothing incomplete, and nothing in vain" (Barnes, 1984, Pol. 1, 1256b 20–21). He argued that any substance possesses its end immanently, that is, by virtue of what it is. Accordingly, he distinguished two modes by which teleology ("final causality") is expressed in nature: *externally*, where an end form is imposed on a substance, as in the case of a fabricated artifact, and *internally*, where the end form arises from dispositions internal to that substance, as in the case of a living organism. Thus, a clay pot is shaped and made of a substance that is waterproof to be consistent with its intended use, whereas an acorn is internally disposed to grow into an oak tree unless forces external to the acorn prevent it. In this sense, an acorn has an internal principle of change and rest that makes it capable of achieving this pre-specified mature state. This principle is called its *psychê*.

Although the organism is supported by tendencies that derive from tendencies implicit in the basic elements, Aristotle's account of organism causality is ultimately internalist. Mid-19th century reinterpreters called this causal disposition its "entelechy" —roughly, its internal teleological tendency in the context of vitalism, the idea that living processes exhibit a non-mechanistic compulsion to achieve certain ends. Vitalism is an extreme form of teleological internalism. For vitalists, the organism's tendency to develop, reproduce, and act to achieve specific results is attributed to an ineffable essence or *élan vital* that is teleologically irreducible. Whether an internalist view requires that the source of living teleology is ineffable has been a subject of considerable debate that we will return to shortly.

In contrast, an extreme externalist approach argues that appeals to intrinsic factors merely describe what needs to be explained and offers no causal account of its influence. The origins of externalism can be traced back to Plato's ideas about the origin of the world, in which artifacts and living creatures were viewed as expressions of the idealized Forms imposed on material entities by a divine craftsman or 'Demiurge' described in his *Timaeus*.

The contemporary scientific consensus is that internalist accounts merely beg the question of how end-directed processes work. This preference for externalism is exemplified by the way that the biological sciences have progressively abandoned intrinsic teleological viewpoints over the past few centuries and almost exclusively embrace external causal counterparts. A prominent example of externalism is Ernst Mayr's (1974) comparison of the end-directedness of organism behavior to the behavior of an automated guidance system or a computer program, the only difference being that whereas machines and programs are created by design, life's end-directedness evolved by natural selection. In this respect, both natural selection and artifact fabrication can be conceived as externally imposed.

Similarly influenced by Kant's (1790) account of organism causality and Wiener's (1948) cybernetics, Maturana and Varela (1980) modeled organismal causality as a circular network of chemical processes that succeeds in producing and maintaining a minimal "autopoietic" (literally "self-fabricating") system. Maturana and Varela distinguished two levels of autopoietic analysis: the *operational*, where cause-and-effect phenomena like chemical reactions occur, and the *functional*, which refers to

the symbols with which an external observer describes the co-dependent network of chemical reactions. In line with this distinction, they noted that "purposes or aims are not features of the organisation (…) these notions (…) belong to the domain of descriptions" (1980, p. 85). Accordingly, notions like ends, purposes, and functions remain at the symbolic level and thus teleological causality is treated as epiphenomenal; merely extrinsic description, not an intrinsic property.

2.2 Constitutive vs. descriptive accounts

Descriptive teleological theories are theories that tend to be agnostic about the dynamics of an end-directed process. This allows them to sidestep directly addressing any metaphysical implications. We provide two examples: Pittendrigh's coining of the term "teleonomy" and the theory of natural selection.

Pittendrigh (1958) proposed the neologism "teleonomy" instead of "teleology" to characterize end-directed processes in general, whether exemplified by living organisms or designed devices, such as thermostats or guided torpedoes. Because it merely described a class of behaviors, irrespective of their origin or mechanism, it was eagerly adopted by biologists happy to ignore dealing with philosophical issues. Because of this it remains widely used today.

Perhaps the best-developed example of a descriptive theory is the theory of natural selection. Darwinian theory[2] describes how the struggle for survival, adaptation, and reproduction can account for how some traits are passed on and others are culled relative to some environment. While the theory has proven enormously successful in this respect, it is largely agnostic about the details of the work that organisms do to survive and reproduce. The means by which a particular organism manages to adapt to its environment and pass on its genes and traits to offspring is outside the scope of the theory. For this reason, natural selection has been sometimes analogized to a passive sieve (Walsh et al., 2002; García-Valdecasas & Deacon, 2024). By describing natural selection as "descriptive" we are not denying that it explains how adaptations arise, only that by leaving aside any analysis of the physical work required for a living system to persist, it focuses on formal consequences of evolution but ignores how these are specifically produced.

In contrast, a constitutive theory of organism teleology must account for the specific causal dynamics of end-directed behaviors. This includes, at least, a basic account of the material and energetic processes involved in their production. In other words, it must focus on *how* organisms maintain themselves far-from-equilibrium, replace their damaged parts, and persist despite constant material and energetic turnover. A constitutive account of teleological causality should explain how the adaptive result is produced, not just why these results are likely.

[2] Of course, if Darwinian theory is not considered teleological, the argument that it is a descriptive teleological theory will not apply. However, some authors have interpreted it to be teleological (Lennox, 1993).

2.3 Targeted vs. terminal processes

Many natural processes produce a direction of change that progressively approaches a specific end state. For example, objects thrown into the air are pulled toward the center of the earth by gravity and fall back to earth until they come to rest on the ground. Opposite poles of two magnets will pull each other closer until they touch. Sugar mixed into tea will dissolve and disperse evenly until the concentration is everywhere the same. And frozen food left out to thaw will eventually reach a temperature that is at equilibrium with the surrounding air.

We can describe each of these processes as *terminal* because they involve a spontaneous tendency to change toward a maximum or minimum value of some variable, beyond which no further spontaneous change is possible. The last two examples are particularly relevant because they highlight an important contrast between teleological and thermodynamic processes. Although both are often described as "end-directed" because they tend to change toward a definite point where change ceases, in most other respects, teleological and thermodynamic processes are opposites.

The second law of thermodynamics describes a spontaneous tendency of things to change until they reach an intrinsically stable (terminal) state. Unless impeded from changing state, a physical system of many parts will spontaneously increase in entropy until it reaches its maximum value at equilibrium. In this respect, physical systems are not "attracted" to a terminal state, nor is work required to push things toward that state. A system reaches equilibrium at the point that it has exhausted all available internal free energy. The system may be constantly changing at this point, but the change is no longer in an asymmetric direction. Cessation of asymmetric change is determined by the fact that some variable of the entire system has reached a maximum or minimum possible value.

Because it takes work to cause things to change in ways that are contrary to what would happen if left to change spontaneously, terminal processes are the norm in physics.

Although teleological processes are also characterized by asymmetric change, this is not change toward a terminal state. The state toward which a teleological process is "driven" is arbitrary with respect to what would occur without intervention, and often quite divergent from the terminal state that would otherwise tend to obtain in that context.

In this respect, a teleological causal process is not so much defined by the end state, but by the countervailing trajectory of that change. We are well-acquainted with end-directed causality. Lifting an object off the floor to set it on a higher surface counters the effect of gravity, pulling apart magnets attracted together by magnetism counters the effect of their attraction, and putting on a coat impedes the dissipation of heat. Each of these actions counters some spontaneous tendency to change toward a terminal state. Even if the "target" state of a purposeful agent converges to a terminal state, such as stirring tea to speed up the rate that sugar dissolves, it is teleological because work is required to modify the spontaneous rate at which this would have otherwise occurred.

This is particularly relevant to the core disposition of life. An organism's targeted disposition is to be maintained in a far-from-equilibrium state. To remain in this

unstable condition and avoid termination, a living system must constantly do work. So, a theory of natural teleology must explain how it produces a direction of change that diverges from what would be predicted from the second law of thermodynamics and related natural tendencies[3].

2.4 Normative vs. nonnormative properties

Unlike nonliving processes, living processes must be actively maintained in an intrinsically unstable state. Because of this instability, the persistence of an organism is not guaranteed. It must do constant work to counter the tendency to succumb to the second law of thermodynamics.

Traditionally, teleological processes have been characterized as being produced *for the sake of* achieving some target state. As Woodfield describes it, it is implicit in the logic of teleology that "an event occurred in order that a second event should occur, that is, in order to produce a certain result" (Woodfield, 1976, pp. 15–16). The expressions "for the sake of" and "in order to" that are commonly used to characterize teleological causality are, of course, descriptive qualifiers with counterfactual implications. Such implications follow from the fact that organism teleological behavior is produced to counter this intrinsic instability. Maintaining a specific far-from-equilibrium state also preserves a specific disposition to change. With respect to organism teleology, then, this is a disposition that exists for the sake of preserving this same disposition. Should that target state not be achieved, this particular cause-effect relationship would not obtain, and a different (often terminal) cause-effect relationship would instead.

In this respect, this either-or relationship between potential causal alternatives has a distinctive self-referential character for living organisms. There is a plus/minus value to which state ultimately obtains. This bipolarity of value is a function of the at-risk status of the system of physical relationships that constitutes this disposition. It is a value that is not "assigned" by an external observer or with respect to some abstract principle, but rather by this basic recursive causality of life. In other words, because this self-sustaining disposition has its own existence as a consequence, it can potentially benefit by its effects or be harmed by their absence. It is this vital consequence that confers a normative character to this disposition.

As a result, the normativity of life is not superimposed by external observation. It is an either-or relationship between potential causal alternatives of what is constantly at risk of dissolution that grounds value talk. Our normative categories are hence not projections.

[3] One of our anonymous reviewers wondered whether hurricanes might exhibit a targeted disposition of the kind that is only attributable to life. The answer is negative. Hurricanes exactly exemplify the distinction that we are making, as well as the contrast between the living and non-living. In our view, a targeted process is distinguished from a terminal one in that it involves work that opposes the latter. Hurricanes spontaneously dissipate the very energy gradients that initiate them. So, they do not do any work that opposes the increase of entropy. In fact, in dissipating all their available energy gradients, they *terminate* themselves. An objection similar to this one is also addressed in the last paragraphs of Sect. 5. We are grateful to our reviewer for having raised the hurricanes example.

2.5 General vs. particular causes

When Aristotle describes the disposition of an acorn to develop toward the form of a mature oak tree, only the general form of this end is implied. The specific shapes of the branches and positions of the leaves are not specified and will depend on varying local conditions. Even the most ardent genetic determinist recognizes that only the general form of the organism is prefigured. In this context, the American philosopher C. S. Peirce championed the idea that natural ends must be understood as generals (or types) rather than as particular realizations (or tokens). He wrote: "we must understand by final causation that mode of bringing facts about according to which a general description of result is made to come about, quite irrespective of any compulsion for it to come about in this or that particular way." (1931-35, CP 1.211, 1902). In other words, Peirce saw that the same final cause can be produced in very different means and realized by very different material substrates, although probably not by any indefinite number of them. Inspired by this, it may be argued that a final cause is indirect, and underdetermines the specific material details of its realization. In contrast, an efficient cause is the direct effect of one particular individual event on another particular individual event.

Both the persistence of an organism and its reproduction is multiply realizable. They involve the transfer of form from one substrate to another, often many times. As noted above, the form that is transferred is of a particular far-from-equilibrium state that is organized to propagate this same form. The precise details of the form and new substrate to which it is transferred only matter so long as this multiply realizable disposition is transferred as well. In this respect, the target that characterizes biological teleology is a constrained range of conditions, not any one individual state. Biological teleology thereby involves processes that constrain future conditions. In other words, the target of teleological change is a constraint, not some thing. So, biological teleology can be described as a disposition to produce a constraint. Unlike a mentally represented form, a biologically-constrained outcome is not in any way abstract. What is less clear, however, is how such a constraint-transferring and constraint-preserving process is brought about.

3 Constraint, work, and life

In recent years, the concept of constraint has received increasing attention (e.g., Shannon, 1948; Polanyi, 1968, Kauffman et al., 2008; Deacon, 2012; Hooker, 2013; Moreno & Mossio, 2015). Its utility in analyzing both biological and informational processes is beyond dispute.

We use the term "constraint" in a somewhat broad thermodynamic sense to refer to the reduced degrees of freedom characterizing some physical process. This diverges somewhat from both the colloquial sense of a physical boundary as well as its formal mathematical sense, but is roughly consistent with its use in complex systems theories. We hope to clarify this in the sections that follow.

Thermodynamic constraints can be thought of as boundary conditions influencing the probable and improbable dynamical variations of a physical process. These

boundary conditions can be both extrinsically imposed or intrinsically generated. For an example of the latter, consider the constrained flow of water in a whirlpool as it drains from a tub. This circular form of water movement is not due to some imposed restriction, but to a collective reduction of molecular trajectories, as certain patterns of flow impede or reinforce one another along their way. Such intrinsically generated constraints are neither physical barriers nor the result of work to counter some process. They emerge spontaneously, following the path of least resistance and maximizing the total free energy for the dissipation of some energetic or material gradient.

Constraints are often conceived as structures or processes. This is unfortunate. A biological constraint can just be the reduction or maintenance of some physical parameter. For one thing, it can be the control of the O_2/CO_2 ratio in blood as a result of heart and lung activity. Because constraints are only relational—what results from the internal organization of a living system, or from its interactions with its environment, they are difficult to cash out in terms of structures or processes, and are often misguidingly so.

Understood relationally, constraints provide a way to define form and order without referring to ideal types. Thus, a more organized process is more constrained in its dynamical variability than a less organized one, and a more symmetrical form is more constrained in the number and diversity of its features than a less symmetrical one.

So, although constraints on change are neither material nor energetic properties, they nevertheless can have significant causal relevance. Indeed, thermodynamic work is an expression of the effect of constraint. Thermodynamic work is the result of the constrained release of energy where the form of this constraint channels its effects. But constraints do not make things happen on their own. In his 1968 *Science* essay "Life's irreducible structure", Michael Polanyi eloquently argued the importance of the distinction between work and constraint for elucidating the thermodynamics of life. He pointed out that physical and chemical laws do not distinguish living from mechanical processes. Instead, both phenomena differ in how these laws are constrained. Comparing life to a machine, he reasoned that both processes take place "under control of two distinct principles." These are physical-chemical principles—natural laws—and boundary conditions—or design constraints. Writing in the early days of molecular biology, Polanyi argued that DNA is the source of constraints that provide boundary conditions on the physics and chemistry of life. He refers to these constraints as the "information" that shapes a growing embryo and regulates the living process.

When either a machine or a living organism breaks down, the laws that govern its physical and chemical processes do not change, only the boundary conditions do. In a machine designed to accomplish some task, breakdown of some of its boundary constraints can render it unable to achieve this end. In a living organism, breakdown of some of its physiological boundary constraints can also result in failure to accomplish certain ends. But this failure is more significant in living systems whose primary end is to prevent their own termination. Thus, at an organism's death, a number of physiological constraints break down, and a vast array of physical and chemical processes that were once prevented become active and cause its decomposition.

Understanding this complementary relationship between constraint and energetic dynamics is critical for identifying exactly what needs to be maintained for life to

persist in its far-from-equilibrium condition. During the lifespan of an organism, the material embodiment and the energetic drivers of change are in constant turnover, only its constitutive constraints persist and are mostly protected from modification. Explaining how this is possible within the strictures of the second law of thermodynamics is not simple.

In a simple but profound quip, Stuart Kauffman and colleagues (Kauffman et al., 2008) provided an important clue. They argued that "it takes constraints on the release of energy for work to happen but work for the constraints themselves to come into existence." This is a deceptively simple point, but its implications for understanding self-maintaining systems are critical. It is this reciprocity between the production of work and the production of constraint that enables life to persist in a stable far-from-equilibrium state. With every work cycle, new constraints can be produced, and these new constraints can be available to channel further work.

The second law of thermodynamics can also be understood in constraint terms. It describes a tendency for constraints to spontaneously degrade in any physical transformation. A reduction of constraint is at the same time an increase in degrees of freedom and less constrained, while a system driven away from equilibrium is progressively losing degrees of freedom and becoming more constrained. To be alive is to constantly resist succumbing to the relentless effects of the second law of thermodynamics. Framed in terms of these two facets of physical work, being alive requires the preservation of constraints that channel work to constantly regenerate these most critical constraints. From this perspective, life can be roughly characterized as a constraint on chemical interactions and their constraint production to do self-maintaining, self-propagating work. This feature, in turn, maintains and propagates the constraint on these interactions. Exactly this dependency relationship between this constraint and those produced by underlying chemical interaction is the critical target condition that best characterizes life. In this respect, the possible degradation of these constraints is precisely what is most at risk.

The specific form of these constraints determines the form of work that can be produced. As a result, without the maintenance of constraints that indirectly contribute to the persistence of the organism, there can be no teleological causality.

4 Constraints as generals

Section 3 reframed the logic of form generation in thermodynamic terms, highlighting the distinct contributions of constraint and the release of energy. In this section, we discuss a notoriously difficult problem for any causal theory: the relation between generals and particulars. This has been a philosophical conundrum for millennia with respect to human agency, and it is no less a biological conundrum today.

Generals are usually characterized as abstract relations, theoretical constructs that seem orthogonal to the particularity of objects and events. If physical change is the result of interaction between particular objects or events, how can abstract features like the continuity of a form, transferred from one material substrate to another in many systems, be physically effective?

In Section 2, we reviewed Peirce's framing of final cause in terms of a "general description of result" (1931-35, CP, 1.211). Unfortunately, this intuition provides no causal explanation of how a "general description" can be physically realized. A description is a representation of something. It is an abstract relation that cannot explicitly be the cause of what it describes. So, taking Peirce's claim at face value, we need to understand the terms "description" and "representation" differently than what is normally understood in their mentalistic sense.

To ground the causality of a general description we will reframe it in terms of constraint. To illustrate how an end-directed process involves both representation and constraint, consider the common experience of following a set of instructions to assemble a device or bake a cake. The set of instructions provides a general description of the steps that must be followed to produce the desired result, but only actually producing actions that are *constrained* by those general descriptions can accomplish the intended task. The instruction set provides constraints on the selection of possible efficient actions that are consistent with achieving the desired general result. They specify the details of these movements only to a level sufficient to achieve this general end and leave many other details unspecified.

Similarly, an experienced driver who intends to reach the opposite side of town does not know in advance the exact step-by-step interactions that will be encountered en route. Because of this, he makes impromptu choices about which route to take contingent on adapting to local conditions that facilitate or impede reaching his destination. The selected means are constantly open to change so long as they enable him to realize his intended result. Nor is the intended result (in either case) specified in all its granular details; it is just constrained with respect to few important details. This underdeterminacy of the specific result is the general character of a constraint.

Analogously, a biological end-directed process can be loosely compared to these mentally mediated cases in which the representation of a potential future configuration guides the choice of work likely to bring that configuration into existence. This potential future configuration does not determine all or even most physical details of what needs to be done; it is just a general description. Yet it nevertheless is a critical influence on the achievement of the desired end. But how?

To see this, consider the evolution of a molecule like hemoglobin. Each hemoglobin molecule has an affinity for oxygen, which is extracted from air or water and given up to somatic cells. The specific structure of a hemoglobin molecule and its capacity to hold onto an iron atom enables it to ferry oxygen throughout the body, but we typically assume only this general effect of hemoglobin has been "selected for." Of course, this is not entirely accurate. Natural selection merely eliminates variants that are less consistent with organism maintenance and reproduction, leaving behind those more concordant with these general constraints. The general nature of these constraints is exemplified by the fact that our bodies have retained multiple hemoglobin variants during evolution. This includes two adult variants and multiple fetal hemoglobin variants (as well as two hemoglobin "pseudogenes" that are present but not expressed). Moreover, many invertebrate species use a copper-based oxygen-ferrying molecule called hemocyanin to do what hemoglobin does for vertebrates in their blood. So, the process of natural selection preserves a constrained range of material features consistent with the same "function" (which itself is a constrained

range of processes). In engineering terms, we might say that a constraint specifies the allowable degrees of freedom of the value of some variable; often described as "tolerance." Natural selection likewise determines a level of tolerance that underdetermines which specific physical tokens will conform to it.

Of course, this general disposition must be instantiated in particular physical constraints. Only in this way can a constraint specify the form of physical work and have physical consequences. So, although constraints do no work by themselves, they bias dynamical processes that can do work to produce new constraints in new physical substrates. This is why Peirce argued that a final cause does not directly produce physical change. Constraints only organize efficient causes so that they can collectively bring a constrained target state into existence.

Living systems have evolved to take advantage of the many physical constraints (like the pull of gravity) and sources of energy (like sunlight) in their environment. The combination of these environmental constraints helps to channel the work required to preserve this capacity. This work, in turn, directly or indirectly preserves and regenerates the constraints that ensure its continuity.

Consider the sequence of nucleotides aligned on a DNA molecule within an acorn. This sequence constrains the structure of the proteins produced, which constrains the interactions of the plant's cells, which in turn constrains the general shape of its leaves, and so on for a myriad of other physiological features at all levels of scale. At each stage, constraints from one level that are embodied in one kind of substrate are transferred to another level and substrate using different mechanisms to preserve and regenerate the capacity to do further work. Many internal and external factors will additionally constrain this process whose interactions will determine the specific physical form of the mature oak tree. The particular physical mechanisms that produce these final details and the materials that embody them are both multiply realizable, although again, there are constraints on the tolerable degrees of freedom possible. Because constraints are only limitations on a system's interactions, there is always a range of tolerable variations that are consistent with the persistence of the system itself. Component constraints and particular mechanisms can therefore change without significant effect on the organization.

Thus, only a tiny fraction of the possible molecular configurations corresponds to a living being. At the same time, every living system like an individual oak tree grows in ways that allow for variability and multiplicity of shape and dimensions while embodying the same general constraints.

Of course, this picture just describes what needs to be explained. We need to see how a physically embodied constraint, like the sequence of nucleotides in a DNA molecule, can have an efficient causal influence and be more than a mere "general description of result." And beyond this, we need to identify what sort of physical system can perform the sort of work capable of generating and preserving the constraints that makes this possible.

5 Autogenesis

In this section, we demonstrate how the *molecular equivalent of a general description* of *a result* can be causally efficacious in an empirically testable molecular model system. This model will provide a sort of *proof of principle* that the properties characterizing teleological causality can emerge from constraints that organize material substrates that otherwise lack these properties.

Analyzing the complex dynamical relationship of linked constraints even in the simplest bacterium poses a practically insurmountable obstacle to analysis. Each bacterium is constituted by an intricate network of thousands of interdependent molecular interactions that defy comprehensive analysis. To discern how these component processes collectively make an individual remains well beyond current science. So, rather than trying to reverse engineer such complex systems, we have taken an alternative constructive approach.

Below we describe an extremely simple candidate molecular model that can help to visualize how a constraint can both represent and influence a chemical process. It is modeled on the structure of a simple virus but differs in several important respects from it. Most importantly, it is nonparasitic and capable of self-repair and simple self-reproduction. Since 2006, Deacon has been exploring the likely properties of a molecular thought experiment that, although simpler than a virus, exhibits the basic features of biological teleology introduced in Sect. 2. It can be described as an *autogenic virus* or *autogen*, for short. The details of this model system have been described elsewhere (e.g., Deacon, 2006, 2012, 2013, 2021, Deacon & García-Valdecasas, 2023). Deacon (2012) calls the dynamic that gives rise to the autogen *teleodynamic.* The autogen model is one example of a teleodynamic system. In general terms, a teleodynamic system like the autogen is normative because it is generated *for the sake of* the persistence of its own integrity and the preservation of its general capacity to counter perturbation.

Here, we will only discuss the features of this model that are relevant to teleological causality (for a more detailed account see Deacon, 2021).

A simple virus, like the polio virus, consists of a container or "capsid" shell which is typically made of protein molecules that assemble themselves into facets of a polyhedral structure that encloses an RNA or DNA molecule. When incorporated into a host cell the viral RNA or DNA commandeers the cell's metabolism to make more capsid molecules and more copies of the viral RNA or DNA. These reassemble into replicas of the original virus.

In contrast, a non-parasitic virus would need to use a different and much simpler molecular process to reproduce its parts. Instead of nucleotide replication, a process called *reciprocal catalysis*[4] is proposed as an alternative mechanism for producing new components. Catalysis is a chemical process in which one molecule can vastly enhance the rate of molecular reactions/transformation without itself being modified. A catalyst provides this by lowering chemical reaction thresholds. This increases the chance that certain thermodynamically favorable transformations that would other-

[4] Reciprocal catalysis is also sometimes described as involving a "collectively autocatalytic set" (see Kauffman, 1993 for a more detailed account of the process).

wise be prevented by high reaction thresholds will occur at far higher rates than in the absence of the catalyst.

Reciprocal catalysis occurs when two catalysts, either directly or indirectly, contribute to produce each other. This can involve two or more catalytic steps to reach a complete circle of reactions steps in which each catalyst is produced by another in the set (Kauffman, 1993). The result is a chain reaction that rapidly increases the local concentration of catalysts and other reaction products. The resulting process is self-organizing in the sense that it rapidly skews the local concentration of catalysts and their products. This decreases local entropy at the expense of increasing global entropy.

The autogenic model system links reciprocal catalysis with another self-organizing molecular process called *self-assembly*. Self-assembly is essentially a variant of the process of crystallization. Viral capsids (virus shells) spontaneously self-assemble (as do cell membranes, microtubules, and many other complex molecular structures within cells). The regular geometries and affinities of the molecules constituting crystal lattices and virus capsids cause them to fit together with one another and spontaneously form into crystals, sheets, polyhedrons, or tubes. Processes of self-assembly are self-organizing in the sense of generating local regularities at the expense of the global increase of entropy.

These two self-organizing processes—reciprocal catalysis and self-assembly—are chemically complementary. Reciprocal catalysis produces high locally asymmetric concentrations of a small number of molecular species, while self-assembly requires persistently high local concentrations of a single species of component molecules. Likewise, self-assembly produces constraint on molecular diffusion while reciprocal catalysis requires limited diffusion of interdependent catalysts. In this way, both reciprocal catalysis and self-assembly produce the boundary conditions that support each other (see Fig. 1).

Their complementarity makes these two processes co-productive and co-dependent. In the autogenic virus, reciprocal catalysis limits the reduction of capsid mol-

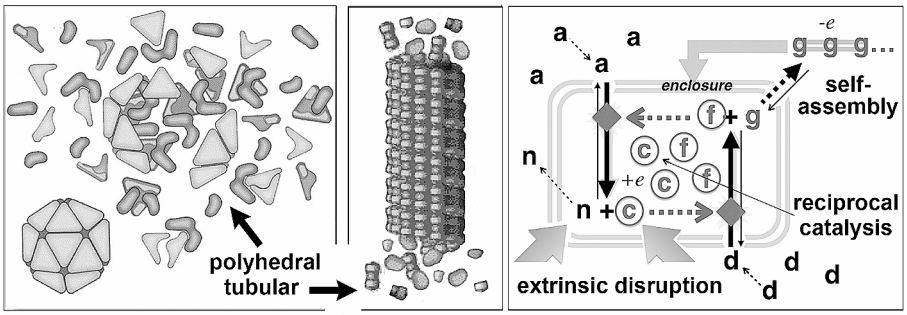

Fig. 1 Depictions of two hypothetical forms of simple autogenetic (i.e., self-reproducing) viruses with polyhedral (left panel) and tubular (middle panel) capsid structure. The chemical logic of simple autogenesis is depicted in the diagram in the right panel. Molecule **a** is modified by catalyst **f** to produce catalyst **c** plus side product **n** and catalyst **c** modifies substrate molecule **d** to produce catalyst **f** and side product **g** which tends to self-assemble onto a capsid and will thereby tend to encapsulate catalysts **c** and **f** and prevent them from diffusing away from one another

ecules due to their accretion to the growing capsid shell, and capsid shell formation limits the diffusion of catalysts that would otherwise decrease the probability of reciprocally catalyzing each other's production.

But autogenesis requires something more than just the reciprocity of these self-organizing processes. It requires co-locality, an additional constraint that results from their mutual dependence. Obviously, this reciprocity only matters if the two processes are somehow co-localized and materially linked. Such a linkage is possible if the two processes share a common element. This can happen if one of the molecular side-products produced by reciprocal catalysis is a molecule that tends to spontaneously self-assemble into a closed capsid structure. At this point, the most rapid and effective capsid formation will tend to occur where reciprocal catalysis is also the most rapid and profligate. This co-localization increases the probability that capsids will tend to grow to enclose a sample of the catalysts that produce each other and produce capsid-forming molecules.

While encapsulated, catalysis will cease, and an inert structure with the potential to be reanimated emerges. Re-animation will occur if the capsid gets damaged and releases its contents into an environment rich in catalytic substrates. Catalysis will then recommence. By producing more catalysts and capsid molecules the damage will be repaired and the inert form will be reconstituted. The system can repeat this cycle again and again, each time modifying and recruiting new molecules from the environment to become part of the autogen (see Fig. 2a). With linkage of these two self-organizing dynamics, each becomes, in effect, the permissive environment enabling the other; effectively becoming a supportive "environment" that "contains" the other. More importantly, this linkage limits each other's spontaneous tendency to proceed to equilibrium—what we described in Sect. 2 as a terminal state. This consequence would be inevitable if these processes were not mutually linked. Reciprocal catalysis would use up substrates and dissipate, and self-assembly would decrease local capsid molecule concentration to the point where no further growth would be possible.

The first law of thermodynamics dictates that none of the material or energy involved in these interactions can be created or destroyed. But the constraints that constitute the autogen can. Irreversible damage can cause the autogenic system to disappear at any time. Should the molecules comprising the catalytic set become so dispersed by disruption that their probability of interaction becomes even modestly improbable, capsid formation would not occur quickly enough to prevent catalyst diffusion and this codependence would be lost. This possibility highlights the critical importance of the synergistic coupling of these linked self-organizing processes. Over and above the chemical constraints produced by each individual self-organizing process, their codependence is a higher-order constraint that depends on the integrity of the whole. It is neither chemical nor molecular, but a constraint on a topological property—i.e. their co-dependent linkage—that remains invariant despite component material changes.

To distinguish this higher-order constraint from the underlying chemical constraints that this constraint links, we describe it as a *hologenic* (or topological) constraint, because it is what transforms these distinct self-organizing dispositions into a unified higher-order dynamical disposition. By preventing the component self-

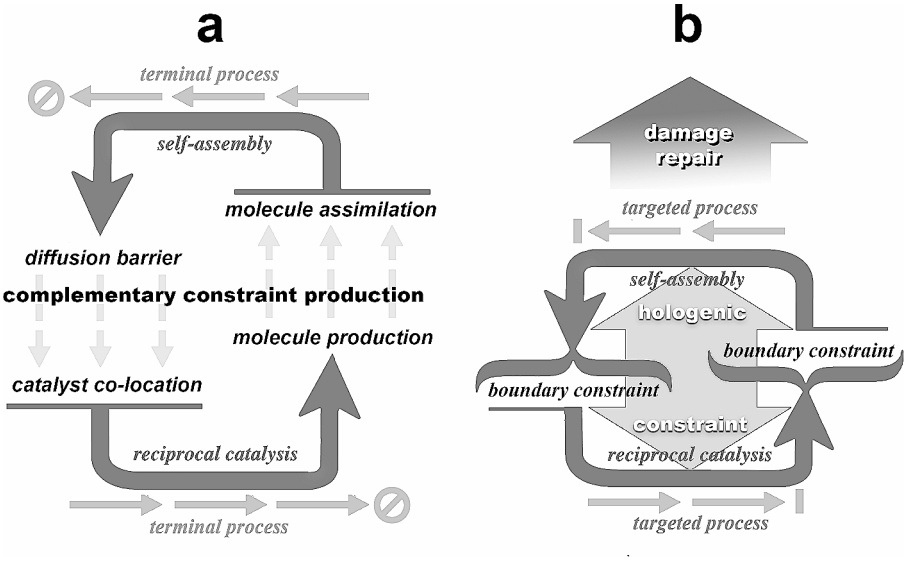

Fig. 2 The left image (**a**) diagrams the reciprocity of the constraints produced by and required for two self-organizing processes: reciprocal catalysis (below) and self-assembly (above). The constraints produced by each and the necessary boundary conditions for each to occur are shown, as well as their complementarities. Both processes are terminal processes (depicted above and below each self-organizing process arrow) that will tend to continue until reaching equilibrium. This will inevitably occur unless they are strongly linked and co-dependent. The right image (**b**) diagrams how these processes are changed in the case that reciprocal catalysis produces a side product that acts as a capsid-forming molecule. This causes the two processes to become strongly coupled and co-dependent. As a result, the component self-organizing processes are prevented from reaching terminal thermodynamic equilibrium and the constraints that are produced are preserved. Although this coupling is due both to a maintenance of proximity and shared material, it is the constraint reciprocity that matters, not any particular material or spatial property. This general constraint is described as *hologenic* (indicated by the large gray arrow) because it determines the integration and unification of all processes into a larger dynamical whole with the emergent properties of self-individuation and self-repair

organizing processes from exhausting their supports and terminating the hologenic constraint (see Fig. 2b).

The hologenic constraint is distinctive in three respects:

First, its multiple realizability is of a more formal character than constraints on its material or energetic substrates. The constraint on the distribution of molecules provided by a balloon or viral capsid, for example, affects specific material substrates. Similarly, in the autogen, constraints on the relative concentration of interacting molecules can affect the rate that reaction products are produced. These are physical constraints on concrete substrates. In contrast, the hologenic constraint is not a constraint on any material or energetic property, but on the *complementarity* among forms of physical constraint. Because this complementarity is a formal relational property, this constraint embodies a higher-order form of multiple realizability different from a balloon or a viral capsid. Also, this complementarity constrains the terminal dispositions of its underlying processes to prevent system termination. Thus, by constraining its own susceptibility to be eliminated, the hologenic constraint indirectly constrains itself.

Second, to use Varela's terminology, the hologenic constraint constitutes a "unity in space and time", that is, a materially individuated disposition. The shared capsid molecule yokes the complementary self-organizing processes to one another to create a localized dynamical-material system that "acts on its own behalf" (Kauffman & Clayton, 2006). In this way, the hologenic constraint is its own beneficiary with an unambiguous self-other distinction.

Third, the way that the hologenic constraint preserves its identity across cycles of damage and repair and embodiment in successive different molecular substrates warrants describing this transfer of constraint as a form of representation. The transfer of a hologenic constraint from substrate to substrate in the damage-repair cycle has its parallel in the genetic information that is inherited in a viral lineage. Though autogenic information is not embodied in a particular molecule like DNA, both viral and autogenic reproduction can be characterized as the transfer constraint. In this way, the autogen meets the basic requirements for minimal evolvability (e.g., Lewontin, 1970, Hull, 1980, Maturana & Varela, 1980, Buss, 1987, Maynard Smith & Szathmary, 1995, Griesemer, 2000, Moreno & Mossio, 2015) though with a very limited capacity for open-ended complexification (see Fig. 3). So long as its synergistic co-dependent integrity (i.e., the hologenic constraint) is not disrupted by irreversible damage, an autogenic lineage may sustain evolutionary modification of both its material con-

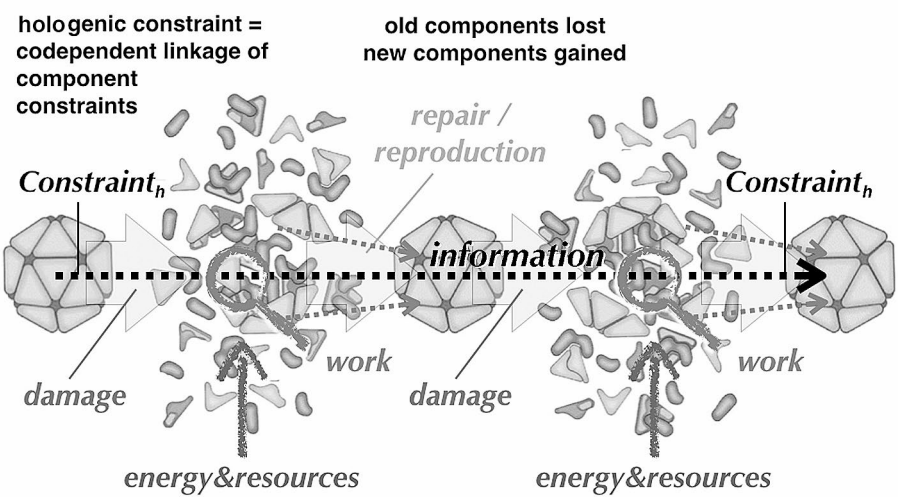

Fig. 3 Two cycles of damage and self-repair in which integrity is temporarily lost but the intrinsic constraints distributed in co-localized molecules enables the recruitment of energy and new substrates from the environment to reconstitute autogenic integrity. As a result, the constraints embodied in the first autogen remain continuously present despite old molecular substrates being replaced by newly synthesized ones. In this way the lineage is delimited by virtue of an unbroken transmission of this hologenic constraint being imposed on and inherited by the organization of future material constituents. The unbroken continuity of this inherited hologenic constraint individuates each inert form and also the lineage, constituting a discrete individual beneficiary of the work that it organizes

stituents and also the details of its component self-organizing processes (see Deacon, 2021 for more details)[5].

These three generic characteristics make the logic of this process generalizable to other forms of target-directed causality in biology[6].

6 From constraint to representation

In the previous section, we demonstrated how co-dependent relations between self-organizing processes can give rise to a basic form of target-directed causation characteristic of living organisms. In this section, we build on these insights to argue that a represented target state can be embodied in the formal-relational features of the hologenic constraint. We conclude that the hologenic constraint embodies a form of minimal representation; possibly the simplest.

In the Introduction, we began by asserting that purposive action has its precursors in non-mental agential modes embodied in the capacity of an organism to respond and adapt to its local environment in a way that supports its continued existence. This is constituted by an organism's capacity to selectively produce work to preserve this same capacity. In the case of an autogenic lineage, the hologenic constraint is the target condition that is preserved intact across this regenerative transition, making it the ultimate beneficiary. This warrants describing the autogenic process as producing and interpreting a minimalistic form of self-representation. In the Introduction, we also mentioned that we use the term "representation" as it is standard in biology, where a DNA molecule can "represent" a sequence of amino acids comprising a protein. This use of the term is neither "mental" nor assumes mentality[7]. In this section, we unpack a naturalistic concept of non-mental representation in the context of the autogenic process.

[5] Deacon (2021) articulates in the details of autogenic evolvability and the emergence of a separate molecular memory system (a genetics).

[6] The general nature of the hologenic constraint is also exemplified by its persistence during the dispersed self-reparative phase of autogenesis, where it constitutes autogenic individuation despite physical fragmentation and uncertain constitution. This drives home the point that the hologenic constraint is not a constraint on specific physical or chemical properties, but on the *relational symmetries* between the constraints that these processes produce.

[7] Although the autogen shares some basic properties with the more familiar teleological experience of mental agency, it also lacks many features of mental representation. Human teleological causality evolved from and depends upon more basic forms of end-directedness but is disproportionally more complex than the normative disposition of the autogen. The evolution of complex nervous systems effectively adds the capacity to impose an internally generated model onto the self-environment relation. Intermediate examples include beehives and bird nests, which realize intrinsic predispositions that are not produced in response to explicit mental images of these results. Only the evolution of the human nervous system with its symbolic capacity makes possible the abstract representation of functional relationships themselves. Based on their mental models, humans can impose functional design onto artifacts that inherit their functional properties. In turn, this form-giving process recapitulates the general logic of basic teleological causality with respect to the external world. So, despite the continuity linking basic end-directed processes and mental agency, the nested dependency of higher-order forms of teleology on lower-order forms shows them to be discontinuous. Providing an adequate account of the evolution of this hierarchic transition will require considerable future theoretical and empirical investigation.

In its most basic form, a representation is an affordance (Gibson, 1977); a property of something that an interpreting system can use to map external relations to internal material or organizational conditions. In this respect, for an autogenic virus to exemplify teleological causality it must both provide the relevant affordance and the means for interpreting it. But unlike a mental representation of an extrinsic condition, the hologenic constraint of an autogenic virus is self-referential. As noted in the previous section, the indirectness of the influence of a hologenic constraint makes it its own affordance. It can only represent an external feature as non-self, for example by initiating the repair of damage.

Many attempts to identify the antecedents of cognition in biology focus on features like metabolism, sensorimotor coordination, and adaptation (e.g., Van Duijn et al. 2006, Godfrey-Smith, 2016, Lyon, 2020, among others). On the other hand, defenders of what is described as the life-mind continuity principle (e.g., as developed in enactivism) assume that life and mind share the same set of basic organizational principles (e.g., Thompson, 2007, Froese & Di Paolo, 2011, and Godfrey-Smith, 2016, among others). In contrast, an autogenic system is neither constantly metabolically active nor does it "enact" its existence by initiating action to modify its environment. But this very simplicity is also its value. It forces us to identify the properties of minimal representation within the small number of its specific physical dispositions.

We argue that three defining properties of representation can be attributed to distinctive autogenic dispositions. These include:

(1) Normativity: the capacity of the autogen to have its own persistence as its target.

(2) Memory: the disposition of the autogenic system to preserve and "remember" its general form of organization.

(3) Discrimination: the autogenic ability to "discriminate" between binary states: inert and active (self-reconstitution).

Taking (1) normativity first, consider the fragile nature of the far-from-equilibrium organization of an autogenic virus, whether in its inert phase or in the process of being reconstituted following damage. While inert, it has a disposition to realize a pre-specified general target state. This disposition embodies the *potential* to initiate chemical work to counter the global tendency for its organization to degrade. And when in the process of reconstituting this organization, chemical work is channeled to counter the ubiquitous tendency for entropy to increase by driving local conditions in the opposite direction. This chemically instantiated disposition thereby has *its own persistence as its target*. In other words, it is its own beneficiary. Notice that although this disposition is a general property of the system, it is embodied at every stage as an individuated physical property, a critical criterion for something to be considered the beneficiary of any end-directed work.

With respect to (2) memory, an autogenic system exhibits, despite structural disruption, a disposition to preserve its general form of organization across changes in material instantiation. Its hologenic constraint is preserved so that its influence can be reimposed in future contexts. To return to a mentalistic example for insight, consider again one's intention to drive to the other side of town. The mental representation of this desired future state is general in the sense that it only involves a description of a type of target state. This description may be no more than a vague mental image with

just enough detail to constrain possible routes to its destination. The vast majority of physical ways this could be realized are irrelevant, only its general form is specified. But as noted above, this abstract form is regularly referred to when deciding what physical steps to take to achieve this target state. It is a constraint on the selection of one from a number of alternative ways of reaching this target state. And yet, as noted above, a constraint does *no* work by itself. It merely constrains the selection of appropriate forms of work.

This same kind of influence is characteristic of the hologenic constraint of an autogenic virus. The hologenic constraint effectively constrains the probability of the occurrence of those chemical processes that are likely to result in reconstitution of a system exhibiting this same disposition. Whereas each complementary self-organizing process produces the physical-chemical boundary constraints supporting the other, the hologenic constraint is physically embodied by that which realizes this co-dependence, and thereby potentiates their future persistence. Like a mental representation, this constraint limits the kinds of work consistent with its physical realization. Unlike a mental representation, it lacks cognition and subjective characteristics. And yet, in both cases, it is the physical embodiment of this constraint that channels the release of energy into a specific form of work that is initiated to achieve a certain consequence. Mentally, this is instantiated in the pattern of neural activity; in the autogenic virus this is instantiated by the catalytic side-products that self-assemble into a capsid container.

The physical embodiment of the hologenic constraint shields autogenesis from the charge of dualism and anthropomorphism, as well as providing the basis for the distinctive causal power of this form of representation. Each component self-organizing process is dependent on the environment for potential energy and molecular substrates to generate its physical-chemical constraints. The hologenic constraint, in contrast, is not dependent on any particular external source. By preserving the co-location and complementarity that supports the continuity of these processes, it re-presents its possibility of being regenerated across potential future cycles of breakup and reconstitution (see Fig. 3) irrespective of physical particulars. In each cycle, new molecules are recruited into its self-organizing dynamics, while the synergistic integrity of its system of constraints is preserved without a break. So long as this co-dependent synergy is kept intact, the component physical-chemical constraints and the whole processes of autogenesis are preserved, enabling the hologenic constraint to embody a memory disposition that preserves the form of the dynamical organization.

The autogenic capacity for (3) discrimination is exemplified by the way it preserves its autogenic capacity despite the constant risk of permanent dissolution. Throughout the entire autogenic cycle, two binary states can be distinguished with respect to one another; one marked—disruption—and one unmarked—inert. The autogenic system exhibits a disposition to discriminate between these two binary states. For the system to critically discriminate between them it must have a higher-order capacity to compare them. And this is not just another system state. Rather, the hologenic constraint provides the reference frame with respect to which the dynamical activity of self-repair effectively "interprets" this binary difference, as well as the *potential* for being in error. Its disposition to initiate work to counter disruption after breakup indicates that the autogenic system is not indifferent to this risk to its continued existence.

There are two obvious classes of potential objections to this analysis.

The first is that our use of cognitive terminology is not merely minimalistic but metaphoric. To this, it can be countered that our use of terms like "representation" goes beyond just simplifying cognitive concepts. We are envisaging the possibility of a natural form of representation as embodied by a physical disposition. To the extent that the initiation of self-reparative work correlates with a critical change in the autogen-environment relationship, the form of this extrinsic change is in fact *re*-presented in the complementary form of this change. So, in this respect, the use of the term "re-presentation" illustrates the way in which this natural form of interpretation is literal.

A second objection might contend that if the autogen can be literally considered representational, why not many other self-organizing processes like crystal formation, where a transfer of form obtains? There is a critical difference between crystal formation and indeed any other self-organizing process, and the autogen model. As discussed above, simple self-organizing processes are terminally disposed. They are not organized to counter disruptive influences. They may tend to reconverge to a prior attractor after the disruptive influence is removed, but they do not initiate work to counter this disruptive influence for lack of a hologenic constraint.

As extrinsic observers, we can discern that self-organization produces the reproduction of a pattern, such as the generation of similarly formed cells in a crystal lattice (and self-assembly) or similar molecules in a chemical chain reaction (and reciprocal catalysis). But while the production of these separate individual structures resembles one another, they only *represent* one another from this external interpretive perspective; in effect, there is no interpretation in these molecular interactions.

In contrast, the hologenic constraint that is intrinsic to each inert individual within an autogenic lineage is not merely similar to its forebears, it was regenerated *by* them. This re-presentation is intrinsically generated. In contrast to the non-normative quality of simple self-organizing processes, its intrinsic re-presentation of a target state makes autogenic dynamics normative.

7 Comparison with competing models

How does the autogen compare to other existing models that purport to explain the emergence of life and its minimal form of teleology? In this section, we discuss the differences between autogenesis and the three most popular paradigms for characterizing the nonlife/life distinction and are thereby treated as potential contenders to explain the teleology of life. These can be loosely categorized as replication-based, self-organization-based, and autonomy-based theories, respectively.

Let us look at them separately.

Replication-based theories assume that molecules that somehow (directly or indirectly) influence the generation of replicas of themselves constitute the sufficient dynamics to be classed with living versus non-living processes. Replication-based theories tend to be presented in gene-centered theories of evolution and in molecular biology contexts. The most widely cited replication-based model systems are RNA-World and protocell theories. These approaches consider reproduction to be the fundamental defining property of life (e.g., Dawkins, 1976, 1982; Jacob, 1993;

Szathmary & Maynard Smith, 1997, Godfrey-Smith, 2000, Hull et al., 2001, Nanay, 2002, Haig, 2020). Because RNA molecules can be carriers of information as well as weak catalysts, it is generally assumed that RNA molecules can, perhaps collectively, replicate new RNA molecules with identical structure. Although this has not been demonstrated and many doubt the possibility, for the argument's sake, let us assume it is possible. If it is, would this cross the threshold from chemistry to life, or from mere physical process to normative process?

Consider normativity. On what basis can we identify replicative error in such a process? Error due to the production of a non-identical partial replica is discernable from the point of view of an external observer. However, because the theory does not provide any corrective mechanism or reparative process (as in autogenesis), the replicator model has no intrinsic mechanism for comparison or determination of replicative error. Nor does replication provide a basis for a self/non-self distinction, a "preferred" target state, or an individual beneficiary; just the continuity of the lineage of similar molecular products. For these reasons, theories that focus on replication as the primary feature tend to treat teleological causality as epiphenomenal.

Self-organization-based theories focus on dynamical processes that produce local entropy decrease and increase order and regularity. The specific chemistry that instantiates it is not relevant, and indeed, non-chemical systems are often treated as relevant exemplars of what distinguishes life from nonlife. Unlike replication-based approaches, form generation and the tendency to reconstitute an organized state after perturbation are taken to be the critical attributes. For example, a recent non-molecular model system called "bio-analogue dissipative structure" (see Kondepudi et al., 2020) describes a collection of metal balls in liquid that organizes into regular branching configurations when a current is passed through them. Although they dissociate in the absence of a current, and their branched organization can be mechanically disrupted, they tend to reassociate into the same or a similar branching patterns when a current is restored.

This dynamical logic has also been studied mathematically (e.g. England, 2013, 2020) and shown to be generative in ways that resembles evolution. England and colleagues have demonstrated that self-organized dissipative systems not only persist in far-from-equilibrium contexts but can also induce the formation of additional self-organized dynamical processes.

The most widely discussed theoretical self-organized model system is the *hypercycle* (e.g. Eigen & Schuster, 1977, 1982). This model describes a collection of self-organized processes circularly linked so that each contributes a critical resource to another member of the collection while each member of the collection is supported by another. In many respects, autogenesis looks like a minimal hypercycle. But there are two critical differences. First, the component self-organized processes in a hypercycle merely contribute support to one another, yet they do not prevent termination. Second, there is no hologenic constraint to establish and maintain their co-dependence. Over the years, critical analyses of hypercycle physics have demonstrated that hypercycles are highly fragile and subject to parasitic short circuits.

In general, self-organization-based approaches are mostly confined to discussions in physics and leave no place for discussing the concepts of self, autonomy, or teleology.

Autonomy-based theories include a range of abstract accounts that trace their roots to Kant's *Critique of Teleological Judgment* (1790/1987). In this classic analysis, Kant distinguishes organisms from machines because of the way that living organisms reciprocally produce each of their parts through the action of other parts of the system. This has given rise to a diverse family of abstract models described as autopoietic (i.e. "self-producing") theories after Maturana and Varela (1980). Examples of autonomy-based theories (e.g. Thompson, 2007, Moreno & Mossio, 2015, Mossio & Bich, 2017) tend to consider the circular production of all components and constraints as sufficient to determine teleological causality, while downplaying both reproduction and the problem of explaining molecular information (representation).

Although autogenesis shares features with these models, it departs in fundamental ways from them. First, a basic autogenic system is more like a virus than a cell, even though it is not parasitic. This difference marks a critical distinction with autopoiesis. Autopoietic theories assume that cellular properties are essential. We argue, however, that cell-based models merely assume system integrity and containment, and implicitly define system unity materially (e.g., by cellular containment) or by assuming that the reciprocal generation of their components and constraints is sufficient to produce individuated unity.

In contrast to such paradigms, the viral analogy explicitly addresses essential properties that each overlooks or merely assumes. Viruses occupy an ambiguous domain intermediate between chemistry and life, marking the boundary region between simple chemistry and what might be described as normative or target-directed chemistry. Viruses exhibit many critical features that we have identified with minimal teleological dynamics, including the transmission of information, evolvability, the preservation and reproduction of individuated unit-selves, and the capacity to be benefited or harmed—as parasitic viruses are. This is the transitional zone where we should expect to find the emergence of the most minimal precursor to teleological causality. So, although an autogenic virus is not "alive" in the sense of a continuously metabolizing cell, it exhibits properties that are fundamental to biology.

A serendipitous advantage of focusing on a model system as simple as a virus is that it avoids a different sort of ambiguity: determining which of the many properties is most relevant. Precisely because autogenesis is so simple, and yet produces such a distinctive end-directed dynamic, there will be a little ambiguity about which features of its organization are most relevant to its core disposition. It is among these few unambiguous chemical and dynamical features of the autogenic process that we need to look to identify a distinct locus of benefit and an antecedently present potential to constrain the initiation of work that enables it.

8 Conclusion

The molecular model described by autogenesis is far simpler than even a virus, but we believe that it satisfies the five criteria for teleological causality outlined in Sect. 1.

First, its target-directed disposition is not reducible to *external* factors, but is a consequence of the holistic constraint on its component self-organizing processes that channels work to maintain its discrete individuality. Second, its target-directed

disposition is *constitutive*, because the linkage between the reciprocal constraint-generating self-organizing processes is physically instantiated by the sharing of a molecule that makes these processes co-dependent. Third, it is *target-directed*, because it is disposed to reconstruct or reproduce only its specific complex of constraints, and because it prevents its component self-organizing processes from reaching their terminal states. Fourth, it is *normative*, because it embodies a disposition to maintain and preserve the causal capacity of which it is its own beneficiary. And fifth, it is a concrete *general*. In favorable environmental conditions, its hologenic constraint can impose what might be considered the general description of an end onto new physical substrates.

More specifically, we have shown how complementary self-organizing processes can become reciprocally linked so that they collectively provide each other's extrinsic boundary conditions *internally*. Because of the way that constraints on dynamical processes can channel work to produce new constraints, they can indirectly re-produce themselves in new substrates. In this way, a hologenic constraint maintaining the reciprocal dependence between physical-chemical constraints can indirectly provide a critical affordance for its own preservation and propagation. As a higher-order formal constraint on the reciprocity of the processes that produce its physical-chemical components, the hologenic constraint ultimately realizes organism individuality. This capacity to indirectly contribute to its own replication in new substrates is multiply realizable. Each new system instantiation of its constraints can be taken to "re-present" the general character of a "type" of material organization. Consistent with the standard of usage in molecular biology, a constraint that is successively transferred to and imposed upon the organization of new substrates conveys "information" in the same sense as do genes.

Based on these criteria, we argue that the autogenic model provides a sort of proof of principle, demonstrating that teleological causality can be materially instantiated, physically efficacious, and not merely epiphenomenal. Unlike the family of Darwinian-inspired theories that redefine teleology as a "selected effect" not a cause, or field theories that attribute all end-directed dispositions (whether targeted or terminal) to extrinsic conditions, the autogenic theory fully naturalizes teleological causality as an intrinsically embodied target-directed disposition.

Finally, we suggest that the autogenic model clarifies the ontological status of both teleological causality and the representational nature of the constraints it realizes. As physically embodied dispositions, teleology and representation have a material existence that can be preserved or lost. Separately, each of the self-organizing processes constituting an autogenic individual are terminal. On their own, they will develop toward a terminal state where their asymmetric dynamics cease, and they no longer exist. Only when they become codependent and the boundary constraints generated by each process facilitate the persistence of the other, is the existence of each of them successfully maintained. The disposition that this creates might be described as *reciprocal termination prevention*. But notice that the expression "termination prevention" might also be a way to describe the essential teleological character of life.

Authors contribution Although some of the underlying ideas were developed by T. W. Deacon in original research published from 2006 on teleodynamics, both authors have contributed to the study conception

and design, to the material preparation and analysis, and to the initial and final draft. Both authors read and approved the final manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This article's research was made possible by funding from two sources: the "Normativity and the Origin of Mind" project (PID2022-140659NB-I00), a project funded by the State Research Agency of the Spanish Ministry of Science and Innovation, and the University of California, Berkeley.

Data availability Not applicable.

Declarations

Ethical approval Not applicable.

Informed consent Not applicable.

Conflict of interest the authors declare not to have any conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Allen, C., & Neal, J. Teleological Notions in Biology, The Stanford Encyclopedia of Philosophy (Spring 2020 Edition), Zalta, E. N. (Ed.). https://plato.stanford.edu/archives/spr2020/entries/teleology-biology/

Barnes, J. (1984). *The complete works of Aristotle*. Princeton University Press.

Bedau, M. (1991). Can biological teleology by naturalized? *The Journal of Philosophy*, *88*(11), 647–655.

Bedau, M. (1992). Where's the good in Teleology. *Philosophy and Phenomenological Research*, *52*(4), 781–806.

Buss, L. (1987). *The evolution of individuality*. Princeton University Press.

Darwin, C. (1959). *On the origin of species*. John Murray.

Darwin, C. (1984). *The various contrivances by which orchids are fertilized by insects*. Chicago University Press.

Dawkins, R. (1976). *The selfish gene*. Oxford University Press.

Dawkins, R. (1982). *The extended phenotype*. Freeman.

Deacon, T. W. (2006). Reciprocal linkage between self-organizing processes is sufficient for self-reproduction and evolvability. Biological Theory 1.2 (2006): 136–149.

Deacon, T. W. (2012). *Incomplete nature. How mind emerged from Matter*. W.W. Norton.

Deacon, T. W. (2021). How molecules became signs. *Biosemiotics*, *14.3*, 537–559. https://doi.org/10.1007/s12304-021-09453-9

Deacon, T. W., & Cashman, T. (2013). Teleology versus mechanism in Biology: Beyond Self-Organization. In A. C. Scarfe, & B. G. Henning (Eds.), *Beyond mechanism: Putting Life Back into Biology* (pp. 287–308). Rowman and Littlefield.

Deacon, T. W., & García-Valdecasas, M. (2023). A thermodynamic basis for teleological causality. *Phil Trans R Soc A*, *381*, 20220282. https://doi.org/10.1098/rsta.2022.0282

Eigen, M., & Schuster, P. (1977). The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften*, *64*(11), 541–565.
Eigen, M., & Schuster, P. (1982). Stages of emerging life—five principles of early organization. *Journal of Molecular Evolution*, *19*(1), 47–61.
England, J. L. (2013). Statistical physics of self-replication. The Journal of chemical physics 139 (12).
England, J. L. (2020). *Every life is on fire: How Thermodynamics explains the origins of living things*. Basic Books.
Froese, T., & Di Paolo, E. (2011). The enactive approach: Theoretical sketches from cell to society. *Pragmatics and Cognition*, *191*(1), 1–36.
García-Valdecasas, M., & Deacon, T. W. (2024). Biological functions are causes, not effects: A critique of selected effects theories. *Studies in History and Philosophy of Science*, *103*, 20–28. https://doi.org/10.1016/j.shpsa.2023.11.002
Gibson, J. J. (1977). The concept of affordances. Perceiving, acting, and knowing, 1.
Godfrey-Smith, P. (2000). The replicator in retrospect. *Biology and Philosophy*, *15*, 403–423.
Godfrey-Smith, P. (2016). Individuality, subjectivity, and minimal cognition. *Biology & Philosophy*, *31*, 775–796.
Griesemer, J. (2000). The units of evolutionary transition. *Selection*, *1*, 67–80. https://doi.org/10.1556/Select.1.2000.1-3.7)
Hacker, P. M. S. (2007). *Human nature: The categorial framework*. Blackwell.
Haig, D. (2020). *From Darwin to Derrida*. MIT Press.
Hooker, C. (2013). On the import of constraints in complex dynamical systems. *Foundations of Science*, *18*, 4: 757–780.
Hull, D. L. (1980). Individuality and selection. *Annual Review of Ecology and Systematics*, *11*, 311–332.
Hull, D. L., Langman, R. E., & Glenn, S. S. (2001). A general account of selection: Biology, immunology, and behavior. *Behavioral and Brain Sciences*, *24*, 511–528.
Jacob, F. (1993). The replicon: Thirty years later. *Cold Spring Harbor symposia on Quantitative Biology*, *58*, 383–387.
Jacobs, J. (1986). Teleology and Reduction in Biology. *Biology and Philosophy*, *1*, 389–399.
Jonas, H. (1966). *The phenomenon of life: Toward a philosophical biology*. Harper and Row.
Kant, I. (1987). (1790/ Critique of Judgment, trans. W. S. Pluhar. Indianapolis, IN: Hackett Publishing Company.
Kauffman, S. (1993). *The origins of order: Self-organization and selection in evolution*. Oxford University Press.
Kauffman, S., & Clayton, P. (2006). On emergence, agency, and organization. *Biology and Philosophy*, *21*(4), 501–521.
Kauffman, S., et al. (2008). Propagating organization: An enquiry. *Biology & Philosophy*, *23*(1), 27–45.
Keller, E. F. (2009). Organisms, machines, and thunderstorms: A history of self-organization, part two. Complexity, emergence, and stable attractors. *Historical Studies in the Natural Sciences*, *39*(1), 1–31.
Kondepudi, D. K., De Bari, B., & Dixon, J. A. (2020). Dissipative structures, organisms and evolution. *Entropy*, *22*(11), 1305.
Lennox, J. G. (1993). Darwin was a teleologist. *Biology and Philosophy*, *8*, 409–421.
Lewontin, R. C. (1970). The units of selection. *Annual Review of Ecology and Systematics*, *1*, 1–18.
Lyon, P. (2020). Of what is minimal cognition the half-baked version? *Adaptive Behavior*, *28*(6), 407–424.
Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and Cognition: The realization of the living* (Vol. 42). D. Reidel. Boston Studies in the Philosophy of Science.
Maynard Smith, J., & Szathmary, E. (1995). *Major transitions in evolution*. Oxford University Press.
Mayr, E. (1974). *Teleological and teleonomic, a new analysis. Methodological and historical essays in the natural and social sciences* (pp. 91–117). Springer.
Moreno, A., & Mossio, M. (2015). *Biological autonomy: A philosophical and theoretical Enquiry*. Springer.
Mossio, M., & Bich, L. (2017). What makes biological organisation teleological? *Synthese*, *194*(4), 1089–1114.
Nanay, B. (2002). The return of the replicator: What is philosophically significant in a general account of replication and selection? *Biology and Philosophy*, *17*, 109–121.
Nguyen, A. (2021). A functional naturalism. *Synthese*, *198*(1), 295–313.
Peirce, C. S. (1931-35). *Collected papers of Charles Sanders Peirce*. Harvard University Press.
Pittendrigh, C. S. (1958). Adaptation, natural selection, and behavior. *Behavior and Evolution*, *390*, 416.

Polanyi, M. (1968). Life's Irreducible structure: Live mechanisms and information in DNA are boundary conditions with a sequence of boundaries above them. *Science*, *160.3834*, 1308–1312.
Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423.
Szathmary, E., & Maynard Smith, J. (1997). From replicators to reproducers: The first major transitions leading to life. *Journal of Theoretical Biology*, *187*, 555–571.
Thompson, E. (2007). *Mind in life: Biology, Phenomenology, and the sciences of mind*. Harvard University Press.
Van Duijn, M., Keijzer, F., & Franken, D. (2006). Principles of minimal cognition: Casting cognition as sensorimotor coordination. *Adaptive Behavior*, *14*(2), 157–170.
Walsh, D. M. (2015). *Organisms, Agency and Evolution*. Cambridge University Press.
Walsh, D. M., Lewens, T., & Ariew, A. (2002). The trials of life: Natural selection and random drift. *Philosophy of Science*, *69*(3), 429–446.
Wiener, N. (1948). *Cybernetics, or Control and Communications in the animal and machine*. MIT Press.
Woodfield, W. (1976). *Teleology*. Blackwell.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.