

# On the naturalisation of teleology: self-organisation, autopoiesis and teleodynamics

Miguel García-Valdecasas 

Adaptive Behavior

1–15

© The Author(s) 2021

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1059712321991890

journals.sagepub.com/home/adb



## Abstract

In recent decades, several theories have claimed to explain the teleological causality of organisms as a function of self-organising and self-producing processes. The most widely cited theories of this sort are variations of autopoiesis, originally introduced by Maturana and Varela. More recent modifications of autopoietic theory have focused on system organisation, closure of constraints and autonomy to account for organism teleology. This article argues that the treatment of teleology in autopoiesis and other organisation theories is inconclusive for three reasons: First, non-living self-organising processes like autocatalysis meet the defining features of autopoiesis without being teleological; second, organisational approaches, whether defined in terms of the closure of constraints, self-determination or autonomy, are unable to specify teleological normativity, that is, the individuation of an ultimate beneficiary; third, all self-organised systems produce local order by maximising the throughput of energy and/or material (obeying the maximum entropy production (MEP) principle) and thereby are specifically organised to undermine their own critical boundary conditions. Despite these inadequacies, an alternative approach called teleodynamics accounts for teleology. This theory shows how multiple self-organising processes can be collectively linked so that they counter each other's MEP principle tendencies to become codependent. Teleodynamics embraces – not ignoring – the difficulties of self-organisation, but reinstates teleology as a radical phase transition distinguishing systems embodying an orientation towards their own beneficial ends from those that lack normative character.

## Keywords

Self-organisation, autopoiesis, teleology, teleodynamics

Handling Editor: Matthew Egbert, University of Auckland, New Zealand

For decades, considerations of teleological causality in biology, or the idea that organisms are intrinsically purposive systems, have been a subject of continued controversy. One reason why this has not abated is that attempts to reduce teleology to more fundamental forms of physical causation have failed to convince many of the wisdom of such strategy (e.g. Bedau, 1991, 1992; Deacon, 2012; Deacon & Cashman, 2013; Gilson, 1984; Hacker, 2007; Jacobs, 1986; Jonas, 1966; Kenny, 1988; Thompson, 2007; Koutroufinis, 2013; Walsh, 2015; Nguyen, 2021). In fact, there is a general consensus that attempts to explain teleology using mechanistic, cause-and-effect explanations have largely failed, and that such attempts will not succeed in the future.

Why should we believe that teleology is a defining feature of life? In his *Critique of Judgment*, Kant (1790/1987) noted that an 'organized being' appears to exhibit

intrinsic teleology because every living process serves as both means and end for some other living process, supporting the integrity of the whole. But he concluded that we cannot understand, much less explain, organic causality as resulting from the straightforward action of mechanistic causes where the origin of motion is external. Living systems are intrinsically predisposed to actively achieve certain results – such as self-maintenance, self-repair and self-reproduction – which tend to preserve this predisposition. This is accomplished by activity that counters the spontaneous degenerative

Universidad de Navarra, Pamplona, Spain

## Corresponding author:

Miguel García-Valdecasas, Institute for Culture and Society, Universidad de Navarra, 31009 Pamplona, Spain.

Email: garciaval@unav.es

effect characteristic of the second law of thermodynamics. Moreover, this end-orientation introduces an irreducible normative character to life that cannot be explained by the laws of chemistry. Chemistry is non-normative in that there are no correct-or-incorrect, better-or-worse, successful-or-unsuccessful chemical reactions outside of life. But a living process can certainly fail to achieve a needed effect and an injured organ can certainly malfunction. So chemical principles and laws are ill-suited to explain the complexity of living processes.

Teleology appeals to axiological notions like the good (Bedau, 1992), the benefit of organisms (Deacon & Cashman, 2013), normativity, agency (Barandiaran et al., 2009) and significance (Von Uëxkull, 1982) that are difficult to cash out in day-to-day biological concepts without losing their meaning. Thus, in recent decades, a number of theorists have tried to unravel teleology in the complex processes of self-organisation. Autopoiesis or autonomy-based theories<sup>1</sup> have attempted to derive teleological notions of normativity and significance from the coproducing and self-making features of living organisation. At their core is the theory of autopoiesis, developed by Maturana and Varela in the early 1970s, and some of its more recent developments like autopoietic enactivism and the theory of biological autonomy developed by Moreno, Mossio and others. These theories have cast new light onto living systems. This article critically examines their treatment of teleology to argue that they leave important questions unresolved. The reason may partly be historical. As known, autopoiesis was initially framed as a non- or anti-teleological theory. While later theories inspired by autopoiesis restored the concept of teleology, there is lack of consensus among these theories on what constitutes teleology and the specific role it plays in biology. In a rough categorisation, they tend to describe teleology in three ways: as either implicitly present in the notion of autonomy or self-determination, as the target of functional attributions in biology or as a combination of both. In all cases, the concept is threatened by reduction either to self-organisation dynamics or to a merely heuristic regulative principle.

Following Aristotle's causal theory, who described teleology as the end for the sake of which something is done – an illuminating principle that was lost in modernity – I will argue that teleology is a causal principle whose primary role is the realisation of some specific good for an organism. Because the rise of end-directed behaviour involves a fundamental change in the dynamics of nature, I will argue that teleology is marked by the emergence of three essential characteristics: (a) individuality, (b) agency, or the capacity of a system to act in its behalf and (c) the capacity to act for a beneficiary. If a biological system exhibits these characteristics, there is a final cause of operation

in it. If it does not, there might be simply far-from-equilibrium, or self-organising dynamics, or the random emergence of structures that appear to exhibit some of the above characteristics without being teleological.

While most autonomy theories see individuality and agency as key teleological notions, and have stressed the primary role of this individual – for example, autopoiesis, the need to constrain the thermodynamic work – for example, the organisational account (OA), work for a beneficiary has less commonly been viewed as part of teleology. In recent years, successful work has shown (e.g. Deacon, 2012; Deacon & Cashman, 2013, 2016) how to operationalise this in far-from-equilibrium dynamics, where the interaction of components is unpredictable and non-linear. While far-from-equilibrium dynamics is not teleological per se, I argue that end-directed behaviour arises as a radical phase transition in far-from-equilibrium dynamics. This transition distinguishes systems whose causal organisation lacks normative character and direction, such as dissipative systems, from systems embodying an orientation towards intrinsically beneficial ends in the context set by the principle of maximum entropy production (MEP). Few autonomy theories have suggested how life can handle the fundamental disposition of MEP tendencies towards self-dissolution and promote the kind of work that prevents it. Only with this phase transition can end-directedness emerge. The end towards which teleodynamics systems are oriented is not merely an end state of self-production, but rather their flourishing despite the ubiquitous threat of degradation. When cast in this way, teleology emerges, and a new and promising picture of life processes arises out of it.

The theory that I will introduce, 'teleodynamics', has been shaped by dynamical systems theory and Kant's account of intrinsic purposes. In the context of the current teleological theories, teleodynamics may be called a 'teleonaturalistic' account (SEP, Allen & Neal, 2020) in the sense that it seeks truth-conditions for biological claims in non-mental facts about organisms. It describes teleology as a distinctive natural cause – a real one, and rejects both teleonomy and other accounts that limit their scope to the explanation of function in biology, such as the main function of the heart, the redness of blood or other biological features whose existence demands explanation (Allen & Neal, 2020). I will thus argue that teleology is a real cause, rather than a mere heuristic principle. To the extent that teleodynamics envisages teleology as a real cause, the teleonaturalistic view of teleodynamics restores final causes to the realm of nature, and moves the conversation from the attribution of function in biology to the kind of causality that teleology is.

In the last analysis, a valid teleological theory must explain the transition from non-life to life, how the

order-reducing tendencies of MEP can be effectively resisted without violation of the second law of thermodynamics and the emergence of an individual beneficiary of thermodynamic work, and provide an empirically testable model of teleological emergence. I think that teleodynamics can successfully provide all these features.

To do so, I will review autopoiesis and the OA. The article is thus divided into four sections. The first section presents the central claims of the theory of autopoiesis, its initial tussle with teleology, Varela's later embrace of intrinsic teleology in a posthumous article with Weber (Weber & Varela, 2002), and as a result of Varela's repositioning, the reason why later autopoietic accounts modified the theory and some developments of which were often at odds with Varela's later views. The second section addresses some flaws in Varela's late understanding of teleology. The third section discusses the so-called OA that was developed to respond to these shortcomings. I will argue that despite their analysis of the limits of autopoiesis, the critical notion of self-determination fails short of providing a full characterisation of biological individuality, and pays no heed to the capacity of living systems to act for a beneficiary. Finally, the fourth section discusses the theory of teleodynamics at length.

## I. Teleology in self-organisation theories

The term 'self-organisation' applies to a wide range of pattern formation processes that result from the local interaction of system components in an initially disordered system. From Kant's notion of purposes to the most recent autonomy theories, there are different ways to construe this notion. For clarity's sake, I will distinguish three of them:

- (a) Kant's natural purposes, which describe the mutual dependence between parts and whole in a living organisation, for example, metabolic processes operate;
- (b) Thermodynamic self-organisation, which describes the rise of systems that spontaneously reduce entropy by constantly changing physical and chemical conditions in the way of dissipative structures, as described by Prigogine;
- (c) Maturana and Varela's autopoietic systems. Their account describes a network of self-regenerating processes within a physical boundary or container that actively respond to perturbations through to maintain system stability.

Let us briefly review them. Kant's (1790/1987) account of life provides the first notion of self-organisation. Kant presciently envisioned organisms as exhibiting an in-built 'formative power' by which its parts reciprocally become both cause and effect of each

other, giving rise to a hitherto unprecedented view of self-organised causality – the landmark characteristic of living beings. Since artefacts lacked any formative power, Kant understood that life's character should not be conflated with that of an artefact. The recursive relation between parts in which no part can be said to be prior to any other could only be explained by what we may call 'endogenous' causality insofar as its origin is internal. This new form of causality stood in contrast with the passive causality of most inanimate matter.

Nevertheless, Kant's account was both limited by his failure to identify the ultimate origin of purposive behaviour and by his reservations about the possibility of a scientific description of it. While we witness and can describe its effects in nature, we remain uncertain as to what exactly causes the characteristic net of internal codependencies that distinguishes life. Kant's explanatory standard, Newton's mechanics, was unable to decipher it. Without being able to explain this mechanistically, Kant left us in the awkward situation of having to make do as if teleology were real while being unable to say exactly why.

Resulting from the pioneering work of W. Ross Ashby, some have seen in complex adaptive systems the realisation of Kant's natural purposes. For Evan Thompson (2007), the complex science of self-organising systems gave 'a detailed scientific characterization of precisely that feature Kant made central to his conception of the organism, namely, a self-producing organization' (p. 138) beyond the set of chemical reactions that inspired Kant's teleology. The science of open systems has shown how processes that seem to run independently interact to create stable, self-maintaining wholes that tend to order and integration. In Thompson's view, this has revealed a new and distinctive form of causality, one that Kant could not even have suspected (Thompson, 2007, p. 139).

The concept of 'self-organisation', though, can be traced back to three different traditions with similar theoretical presuppositions: the feedback circuit model of Wienerian cybernetics, Piaget's general theory of organisation – the first to coin the term 'organisational closure' – and Maturana and Varela's critical revision of their theory. So, the theory of complex adaptive systems, emerged in the 1980s, joined an already crowded field.

In the early 1970s, Maturana and Varela introduced the term 'autopoiesis' (literally, self-production) to describe the self-producing organisation of all forms of life, according to which any organism is a network of component processes that produce themselves. Originally, Maturana and Varela described an autopoietic system as consisting of processes that (a) recursively depend on each other, (b) constitute a unity in their domain and (c) determine a domain of possible interactions with the environment (Varela, 1979, p. 55). The combination of these features involves, on one

hand, that the system is organised as an interlocking network of processes that tends to stabilise or amplify its organisation over time, being ‘operationally closed’. This entails that these processes continuously produce ‘the components that specify [the system] and at the same time realize it’ (Varela, 1997, p. 75). To say that a living system is operationally closed means that it is causally circular (Thompson, 2007, p. 45), so that the net effect is its own production and stabilisation through time within a certain physical boundary. Maturana and Varela singled out the bounded cell as the minimal autopoietic system that exemplifies operational closure, and openly questioned that systems smaller than a cell could exemplify closure.

Maturana and Varela argued that autopoietic systems succeed in maintaining their relational unity, despite being in a permanent flux of chemical reactions. In their view, these processes exhibit some unity, one that gives rise to an invariant organisation. This organisation may be said to be the kind of ‘identity that the living permanently produces and keeps, thanks to a permanent and self-determined change in the physico-chemical components’ (Bich & Damiano, 2008, p. 357; Maturana & Varela, 1973). This identity is the one that is passed on in successive iterations of the same processes, remaining unchanged throughout them. The maintenance of this identity through a permanent flux of chemical reactions was a key premise to understand that the autopoietic organisation is not just a way of describing the mere emergence of structures of greater complexity.

Despite these theories assuming a different view of the concept of ‘self-organisation’, Thompson (2007) stressed that all of them try to make sense of the net of codependencies that produces a self-maintaining whole. And in fact, among these theories, he believes that Maturana and Varela’s model accomplishes what Kant had attempted to describe, furnishing a ‘naturalised, biological account of Kant’s notion of natural purpose’ (p. 140).

So does autopoiesis represent the realisation of Kant’s natural purposes? While the theory may have been a first step towards a theory of natural teleology, autopoiesis and teleology have had a chequered relation. As originally proposed, the theory distinguished two levels of explanation: the *operational*, which maps onto specific chemical or biological processes to reflect cause-and-effect phenomena, and the *functional* or *symbolic*, which refers to the symbolic tools used by an external agent to describe biological processes. Initially, the autopoietic model was entirely built on the operational level. In Maturana and Varela’s words,

purposes or aims are not features of the organisation of any machine (allo- or autopoietic); these notions belong to the domain of our discourse about our actions, that is, they

belong to the domain of descriptions (...) in some encompassing context. (Maturana & Varela, 1980, p. 85)

As a consequence, they deliberately left aside notions such as ends, purposes, goals or functions (Thompson, 2007, p. 147); they are extrinsic to the system and part of the cognitive process of an observer. In this conviction, the theory proved to be Kantian in a literal sense. In the early years, Maturana and Varela (1980) stressed that living systems were no more than ‘autopoietic machines’ that are ‘purposeless’ (p. 86). Although their use of the term ‘machines’ may have been devoid of any functionalistic connotations – and probably meant just ‘mechanical entities’ rather than ‘mechanical artefacts’ – the claim sat uncomfortably with Kant’s rejection of a mechanistic explanation for organisms.

Nevertheless, by the end of the century, Varela changed his mind. In a posthumous article, he backed away from his earlier anti-teleological claims to proclaim that there is ‘intrinsic teleology’ in life (Weber & Varela, 2002, p. 101). Inspired by Jonas (1966) and his organism-centred view of life forms, Weber and Varela (2002) claimed that teleology is the only possible way in which organic life can exist (p. 111) and be fully made sense of. In their view, end-directed behaviour is so intrinsic to biological self-organisations that the concept of causality itself follows from the teleology of the living rather than vice versa (Weber & Varela, 2002, p. 110).

How is Varela’s turn to be interpreted? Di Paolo (2005) thinks that Varela’s later discovery of teleology, presumably inspired by Jonas’ intuitions about the connection between metabolism and teleology, was timely, and that it helps to narrow the gap between biology and sense-making. But in his view, the introduction of teleology marks a radical departure from the original autopoietic model. He specifically suggests that the later Varela may have failed to examine the import of his newer ideas for the operational and functional/symbolic distinction. In fact, there is some consensus that Weber and Varela’s insights lack continuity with the principles and vocabulary of autopoiesis. Di Paolo (2005) suggests that the most valuable views contained in Weber and Varela’s 2002 article are ‘points of reference from which we can provisionally designate the phenomena to be explained’ (p. 432) – that is, the explanatory targets of scientific analysis. Nevertheless, and more in line with the early days of Maturana and Varela, he holds that the correct method to successfully ground critical features like adaptivity or sense-making (Di Paolo, 2005, p. 434) should not use teleological notions, but arrive at them using purely descriptive vocabulary. Other authors (Froese & Stewart, 2010, p. 9) have also pointed to several other weaknesses of autopoiesis.

Other theories influenced by autopoiesis see it as a pivotal step to autonomy theory. Like autopoiesis, autonomy theory stresses the singular causal regime by which biological systems can produce and maintain the conditions that keep these systems in existence. Its main sponsors, Moreno and Mossio, have argued for a naturalised concept of teleology based on the concept of self-determination, which is formalised in terms of constraints. Self-determination conceives self-organisations in terms of open systems that create the conditions for their existence in sync with their environment. By internalising their own constraints, these systems become ‘autonomous’, and by being open to the environment, they are ‘situated’ (Moreno & Mossio, 2015, p. xix) in a way in which the original autopoietic systems were not. In Moreno and Mossio’s view, autonomy and the system’s coupling with its environment on an open-ended evolutionary route is all that is required to make full sense of living phenomena. As Ruiz-Mirazo and Moreno (2012) put it,

(...) the unfolding of autonomous systems and their long-term maintenance depend on their insertion into an open-ended evolutionary route. So it is really the integration of these two main ideas, autonomy and open-ended evolution, that provides a complete, rich enough picture of the phenomenon of life. (p. 27)

Although Moreno and Mossio do not explicitly engage the key autopoietic distinction between the operational and the functional/symbolic level nor comment on Di Paolo’s view, they attempt to bring teleology back to the operational level without any loss of meaning by parsing teleology as a function of biological processes. Their account is largely shaped by MacLaughlin (2001), Christensen and Bickhard (2002) and others who have at length argued that autonomous self-organisation may be sufficient to ground the concepts of normativity and teleology. They hold that teleology is a dimension of functionality (Moreno & Mossio, 2015, p. 36), a function of the system that arises out of its processes. At the same time, teleology works as an explanatory device used to safely argue for the existence of a system on the basis of its self-maintaining effects. So, by the word ‘teleological’, they mean that the effects of the system contribute to explain why the system exists. Their view is that,

On the one hand, the causal regime of a self-maintaining system provides a naturalised grounding for the teleological dimension. Since the activity of the system S contributes, by exerting a constraint on its surroundings, to the maintenance of some of the conditions required for its own existence, the question ‘Why does S exist?’ can be legitimately answered by ‘Because it does Y’. This justifies explaining the existence (again, in the specific sense of its *maintenance* over time) of a system in ‘teleological’ terms

by referring to its causal effects. (Moreno & Mossio, 2015, p. 70)

To summarise, theories inspired by autopoiesis have interpreted teleology in similar ways. First Varela, and Thompson thereafter, have championed an interpretation of self-organising processes as intrinsically teleological and the origin of sense-making. Thompson (2007), in particular, insisted that an autopoietic system, conceived as the totality of interrelated, self-organising means and ends (p. 141), fulfils Kant’s definition of an intrinsic purpose (p. 138). While crediting the late Varela for his restoring teleology, Di Paolo noted the limits of autopoiesis to make sense of the graded response of the organism to changing environmental conditions, and advanced his notion of precariousness. While this notion brought teleology back, Moreno and Mossio have done more to insert this notion in the context of the etiological views of function attribution in biology. On the face of this, theories inspired by autopoiesis agree that teleology is intrinsic, and an autonomy-dependent feature of living systems, and disagree on operational closure, and the role played by terms like ‘significance’ or ‘valence’, which found their way into Weber and Varela’s 2002 account (p. 117). The OA argues that these terms are anthropomorphic and scientifically incongruous (Moreno & Mossio, 2015, p. 101).

## 2. Is autopoiesis really teleological?

I turn now to assess the validity of the teleological claims of autopoiesis. As mentioned, the original autopoietic model confined teleology to the functional or symbolic level. But Maturana and Varela (1980) identified an important point; they stressed that autopoietic systems ‘subordinate all the changes to the maintenance of their own organisation, independently of how profoundly they may be otherwise transformed in the process’ (p. 80). The organisation will maintain itself ‘as long as its basic concatenation of processes is kept intact in the face of perturbations’ (Varela, 1997, p. 76). While the tendency to resist perturbations and stabilise the system may not have passed as teleological, today they are considered a key teleological feature. So, if not already implicit in the claim that an autopoietic system subordinates its changes to the maintenance of its organisation, I will assume that the theory of autopoiesis can integrate Weber and Varela’s (2002) idea that there is a ‘basic purpose’ (p. 117) in the system maintenance. On that assumption, the question is as follows: is the reciprocal network of organisational processes sufficient to account for, or even simply describe teleology?

This question is not a trivial because self-organisation theories describe teleology in contrasting ways. As per Bedau (1992), ‘the primary datum that a

theory of biological teleology must explain is why so many biological phenomena seem teleological' (p. 804), rather than – we may say – restating the fact that these phenomena are self-organised.

In this, it is implicit that any successful teleological theory should account for the existence of systems that are organised to promote specific outcomes. And it is also implicit that these outcomes, by enabling self-maintenance, are self-beneficial. So, can the active response to perturbations in the characteristic way of autopoiesis explain normativity? Consider a bacterium swimming up a sucrose gradient. Presumably, its movement pursues some sort of good or benefit. In a 'value-centred' (Bedau, 1992, p. 781) account, any end-directed behaviour is only real when oriented to some good. This good presumes the existence of a beneficiary, which is the telos of the process in question. More arguments to justify the need for a normative notion of teleology will be presented in Section 3.

So, the following question may be asked: has autopoiesis an ultimate beneficiary? Because Maturana and Varela considered that autopoietic systems were purposeless, they avoided any talk of a possible beneficiary of the reciprocal network of codependencies. At the same time, they argued that autopoiesis exhibits a dynamic identity against a background of chemical components that are not part of its interactive network, and are normally outside its physical boundary. That stability, which is the result of closure, gives rise to an 'emergent or global coherence' (Varela, 1997, p. 73). Varela (1997) describes it as 'a unitary quality, a coherence of some kind' (p. 73) that 'comes about' (p. 74) by the unfolding of autopoiesis. He also described the emergence of a 'virtual self' in a system that has separate local components in which there is no centre, and of which it seems counterintuitive to say that it possesses a real self. As I mentioned in the previous section, this identity was so critical to the autopoietic system that Varela saw in its rise the defining mark of autopoiesis.

Consider autocatalysis, a chemical reaction in which some catalysts increase the production of other components of the set to produce a dramatic increase in the molecules of the autocatalytic set. Consider an autocatalytic set where molecule A catalyses the production of B, which catalyses the production of C, which in turn catalyses the production of A. Could autocatalysis be considered an autopoietic process? Except for the fact that autocatalytic sets lack a physical boundary, they are both self-producing and self-regenerating, and they may describe a causal process that has 'a unitary quality, a coherence of some kind' (Varela, 1997, p. 73) that resists perturbations. And yet, autocatalytic processes do not work for the benefit of a self. While the autocatalytic set spontaneously creates a self-assembling spatiotemporal unit, these resulting spatiotemporal units are neither selves nor the beneficiaries of autocatalysis because the autocatalytic set undermines of its own

basis (Deacon & Cashman, 2013, p. 295). In the cycle described above, the subsequently produced As, Bs or Cs are other molecules different from the originals, rather than beneficiaries of the process. One might even ask whether the beneficiary of autocatalysis is the 'type' of molecule – since more are replicated – or perhaps the so-called 'set'. But of course, the 'type' and the 'set' are just conceptual abstractions. 'With no particular unitary beneficiary – no self for which these processes take place (...) there is no actual teleology' (Deacon & Cashman, 2013, p. 296). Other autopoietic, self-organising systems like snow crystals could be equally teleological.

It is known that Maturana and Varela (1980) denied that autocatalytic processes could be autopoietic 'because (...) they do not determine their topology. Their topology is determined by a container that is part of the specification of the system, but which is independent of the operation of the autocatalysis' (p. 94). At face value, since the autocatalytic set appears to lack a container, this set fails to constitute a dynamic identity of the kind of a bounded cell, in which autopoiesis is usually modelled. But other than for its lack of a physical boundary, it might be asked the following: what prevents us to consider an autocatalytic process or other reciprocal processes in smaller-than-the-cell entities as autopoietic?

Except for the fact that an autopoietic system works within a physical container, Varela's account of closure just concerns process dynamics. It describes processes that (a) recursively depend on each other, (b) constitute a unity in their domain and (c) determine a domain of possible interactions with the environment (Varela, 1979, p. 55). At the same time, it underlined the complementarity between organisational and structural aspects of living systems – their actual material components. And of course, if these components are multiply realisable, that is, if different structures can satisfy the constraints of autopoietic organisation and be made to work, we may wonder why should the physical boundary be held up as an essential ingredient of the model? As per Varela (1991), 'the attribute "living" in the (...) description [of autopoiesis] must address the process that allows such constitution, not the materialities that go into it, or an enumeration of properties' (p. 80). On the face of it, if a physical border may be deemed as part of the 'materialities' of the system, rather than of its basic organisation, why self-organising processes lacking a physical boundary that are nonetheless the result of the 'preferential neighbourhood relations' (1974, p. 193) might not be minimally autopoietic? If so, it could be argued that to exclude autocatalytic processes from the realm of autopoiesis may be somewhat arbitrary.

While Varela took decisive steps to define living systems as self-constituting and self-regenerating, it is unclear why the self-organising dynamics of autopoiesis

cannot be reproduced in smaller self-constituting and self-regenerating systems. From the perspective of the OA, a living organisation is not defined by a physical boundary, but by the way in which the system creates its own conditions of existence (Section 3). Similarly, teleodynamic theory describes the emergence of self-organising processes that stabilise despite the continual threat to degradation posed by the MEP principle (Section 4). Ultimately, both theories appear to better embody Varela's intended project of characterising the living system as a self-maintaining organisation, rather than as a specific molecular composition and its contingent historical configurations (Varela, 1991, p. 83), and both seem to better identify the features that underwrite the emergence of life.

### 3. Organisational accounts

Autopoiesis has been criticised (Di Paolo, 2005) because its model can only cater to the all-or-nothing norm of self-continuance of autonomous systems, rather than to their gradual and restrained steps of environmental adaptation (Thompson, 2007, p. 147), which makes the autopoietic model overly rigid. The model had to be amended to account for the graded reactions of an autonomous system to its environment. In his attempt to pave the way for embodied cognition, Di Paolo (2009) argues that 'a system (...) requires (...) access to how it currently stands against the all-or-nothing barrier given by that norm [the norm of self-construction]' (p. 15). In other words, the system must determine what to do both to prevent dissolution in the face of a potential threat and to benefit from an opportunity. To that end, the structural coupling with the environment seems inadequate because this is blind both to the current status of the system and to what Di Paolo calls 'the virtual consequences of current tendencies' (Di Paolo, 2009, p. 15), that is, what might happen to system integrity if certain conditions were coincidentally met.

To this end, and inspired by the central ideas of autopoiesis, Moreno and Mossio have revised central parts of the autopoietic model for two basic reasons: first, because in continuity with Di Paolo's adaptivity model, the structural coupling of the system with the environment is non-deterministic and hostage to constant change, and second, because the original autopoietic theory failed to consider its fundamental thermodynamic context, ignoring the fact that the system uses energetic and material resources (Moreno & Mossio, 2015, pp. xxvii–xxviii). In contrast, the OA is 'situated' in its environment and grounded in thermodynamics, following previous research into complex systems they hold that autonomous systems can only exist in far-from-equilibrium conditions.

Moreno and Mossio identify the self-determination of a self-organising system – together with its openness to change and evolution – as the feature that underwrites its autonomy. In their perspective,

Biological systems determine (at least in part) themselves, we will contend, by constraining themselves: they generate and maintain a set of structures acting as constraints which, by harnessing and channelling the processes and reactions occurring in the system, contribute to sustain each other, and then the system itself. (Moreno & Mossio, 2015, p. xxix)

They hold that organisms may be characterised as systems that maintain or preserve their system of constraints, rather than systems that generate or create themselves as wholes (Moreno & Mossio, 2015, pp. 5–6) by using their components. This already marks a departure from Kant's notion of self-organisation. While an organism uses components generated by itself, it is a system of constraint preservation rather than of component preservation. In fact, it channels the flow of energy towards its own maintenance in a singular way, which Moreno and Mossio describe as 'collective self-maintenance'. This feature is presented in this way: in biological systems, 'constraints are not able to achieve self-maintenance individually or locally: each of them exists insofar as it contributes to maintaining the whole organisation of constraints that, in turn, maintains (at least some of) its own boundary conditions' (Moreno & Mossio, 2015, p. 17). Biological systems are said to differ from mere dissipative systems in that, rather than individually or locally enabling system stability, their constraints realise closure collectively, that is, in their coordinated and reciprocal work. Each constraint does not 'determine itself' individually, as they say, but enables the existence of one or many of the other constraints. In this way, these can only exist by supporting the whole system – the only way a constraint indirectly supports itself. As a result, the mutually supportive constraint network becomes less dependent on exterior factors, and becomes more internally resistant to threats and perturbations.

Collective constraint marks the distinction between self-organisation and closure. The step is described as the 'takeover of (some of) the boundary conditions required for the maintenance of the system' (Moreno & Mossio, 2015, p. 17). To sum it up,

the higher degree of complexity inherent to autonomous systems in comparison with self-organising ones corresponds to a higher degree of self-determination, because of the takeover of boundary conditions over which dissipative structures have no influence or control. The qualitative change from minimal (self-organisation) to collective (closure) self-determination goes hand in hand, then, with

a quantitative increase of the underlying complexity. (Moreno & Mossio, 2015, p. 18)

What about teleology? I earlier mentioned that Moreno and Mossio attempt to bring teleology back to the operational level without any loss of meaning. Inspired by MacLaughlin (2001), Christensen and Bickhard (2002), autonomous self-organisation suffices to ground the concepts of normativity and teleology. Analysing the role of these concepts in the current debate on functions, they reject the so-called theory of ‘selected effects’, which they find reductive and inadequate to account for functional traits, but accept the basic intuition of the etiological view about the purpose of these traits. In their description,

an organisational account (OA) is the idea that functional ascriptions do account at the same time for both the existence of functional traits and their current contribution to a system capacity, since functions make sense only in relation to the specific kind of organisation which is characteristically at work in biological organisms. (Moreno & Mossio, 2015, p. 70)

To be sure, functional ascriptions in biology are typically defined in terms of the persistence of the containing system (organism or lineage) because they typically bolster it. To the extent that the preservation of favourable traits bolsters the organism or its lineage, such ascriptions are teleological, and may also be either dependent or emergent from the underlying constraints that enable the preservation of such traits.

Call ‘S’ a self-organising system. Moreno and Mossio rephrase the question ‘Why does S exist?’ by noting ‘because it does Y’ and Y promotes the existence of S. This answer stipulates that S’s doing Y is both a reason for S’s existence and fulfils a function that benefits S. Thus, in their view, a self-organised system shores up its own existence through its given functions. In a more recent article, Mossio and Bich elaborate the following:

What makes biological organisation teleological? The core of our argument consists in establishing a connection between organisation and teleology through the concept of self-determination. Biological organisation determines itself in the sense that the effects of its activity contribute to establish and maintain its own conditions of existence: in slogan form, biological systems are what they do. (Mossio & Bich, 2017, p. 1090)

To put it more simply, the OA describes the rise of systems that enable and promote their own conditions of existence through collective self-maintenance (Mossio & Bich, 2017, p. 1090).

The OA has several advantages over other theories influenced by autopoiesis. First, it considers the role of far-from-equilibrium thermodynamics in biological systems as more than mere dissipative structures. Second, the account provides a better answer to the problem of the identity and basic action of a biological system (see Moreno & Mossio, 2015, ch. 4). Third, the OA proves faithful to its purported intention of explaining the teleology of biological systems without using axiological vocabulary, just by looking at teleology as a function of the emergent causal regime brought about by self-determination that is teleologically oriented to its self-maintenance. These joint features make the OA a more elaborate realisation of the self-constituting and self-regenerating dynamics of autopoiesis.

Does the OA specifically address the value-centred question of teleology, that is, of the beneficiary of the closure of constraints? We know that ‘biological individuality (...) has much to do with organisational closure, to the extent that one may conjecture that closure in fact defines biological individuality’ (Moreno & Mossio, 2015, p. 23). But what is the relation between the closure of constraints and the individual? Is this individual a beneficiary of organisational closure? How is closure organised for the sake of this individual? While the assumption may be that constraint closure is there for a beneficiary, the OA does not explicitly say it, and the idea is never specifically argued for.

For this reason, there seems to be room to say that while identifying fundamental teleological dimensions, the OA could benefit from a genuine teleological perspective.

In the remaining of this section, I will describe three ways in which the teleological account of the OA could be teleologically enhanced: (a) biological individuality may be better determined, (b) thermodynamic work must involve a beneficiary and (c) the critical difference between a conditional and a teleological explanation should be noted.

Let us address these criticisms in order:

- a. Biological individuality may be better determined. While the OA correctly envisages the non-reproducibility of the closure of constraints in self-organising entities, the following question may be asked: does the closure of constraints actually constitute a biological unity? A biological individual is not a mere collection of constraints. In the fully determined individual, all its constraints should converge into an integrated unity. Individuals belong to an identifiable class and normally behave in the way in which the members of this class do. On the face of it, just to be a system of reciprocal constraints may not guarantee individuality. While it is true that in a biological individual, every constraint supports its own existence by supporting the existence of other constraints, this merely describes



the way in which a self-supporting network of processes increases its relational complexity to become an organisationally differentiated system. The question is whether this reciprocal network constitutes a unified class. Of the possible ways in which this question may be addressed, I will simply focus on the importance of the unity of the constraint network. It may be argued that biological identities make up classes because the constraints that constitute them have themselves been constrained by a higher order constraint. This higher order constraint is the reason why the constraint network gets unified and resists its disintegration. If this is true, the emergence of a biological individual may require more than self-determination: it requires a second-order constraint sitting above the constraint system whose goal is to prevent its own dissolution. Only when closure is viewed as the outcome of a second-order constraint, can we accurately say that the constraint network makes up a single and unified set. And so, the transition from self-organisation to self-determination envisaged by the OA should be marked not only by the rise of a new causal regime but also by the critical changes that turn a collection of self-organising thermodynamic processes into a unified set. To the extent that the OA does not posit the existence of this higher order constraint, it might be reasonably argued that biological individuality is underdetermined – and in fact, Garson (2019, p. 56) made a similar argument when reviewing Mossio's OA.

Biological individuality is a critical issue because only an individual that acts to preserve its individuation can be the beneficiary of closure, or of any other system behaviour. To more precisely elaborate how second-order constraints arise, Section 4 will briefly discuss the autogen model, and give reasons why teleology may be better captured by it.

- b. Thermodynamic work must involve a beneficiary. I earlier argued that teleology is rooted in the good of the system and its beneficiary. Thus, any specific feature of a system has to be self-beneficial or do something that is self-preserving, even if any process may also involve other external beneficiaries. Hence, beneficiaries become the reason why biological functions exist in the first place. This is how Aristotle and others understood teleology. For instance, Kenny (1988) and Hacker (2007) argue that the biological good of a living being is inherently linked to the system's welfare. To wit, '[a] thing (organ or artefact) has a function only if it exists for the sake of a good' (Hacker, 2007, p. 167), or again, every process in a living system is 'subservient to the welfare of the agent' (Hacker, 2007, p. 179). Comparable or similar arguments

have also been developed by Bedau (1992) and MacLaughlin (2001) and are not entirely novel. If this is true, work for some good, or for some beneficiary, is a necessary criterion for determining whether a process is indeed teleological.

Recall that teleology describes the behaviour of systems that act for the sake of a beneficiary. Does the closure of constraints emerge for the sake of a biological individual or its benefit? Does the OA assume that in living systems, the thermodynamic work is done for a beneficiary? While Moreno and Mossio do not intend to make teleology a merely functional description and reject the reductive view of teleological theories of function, it appears that the so-called 'orientation to some good' of living systems is not provided by the closure of constraints and seems posited as a property of the system by assumption.

- c. The critical difference between a conditional and a teleological explanation should be noted. One key point of the OA is the particular role assigned to functional ascriptions. Far from being mere descriptions of specific traits of organisms, functional ascriptions 'provide an understanding of some of its essential properties and activities' (Moreno & Mossio, 2015, p. 63). In a general sense, functional analyses in biology are correct if (a) they are the main effect of an organ – for example, the distribution of oxygen and nutrients in the bloodstream with respect to the heart – and (b) they contribute to the self-maintenance of the organism or its reproduction. Moreno and Mossio argue that 'functional attributions (...) introduce a teleological dimension into the structure of explanation, in the sense that the existence of a trait could be explained by appealing to some specific effects or consequences of its own activity' (Moreno & Mossio, 2015, p. 63). This might be, however, a more contentious claim for the two following reasons.

First, as argued, teleological claims obtain when they capture a feature that is good with respect to a beneficiary; second, in the absence of such a good or its beneficiary, we run the risk of rolling a teleological explanation into what may be called a conditional explanation, without minding the difference between the two. Let us say that a conditional explanation focuses on A's doing X as being a necessary condition of B's doing Y, and consequently, as establishing some necessary link between two phenomena, in which B's doing Y is the effect of A's doing X in an etiological sense. According to Hacker,

That A's doing such-and-such is a necessary condition of the possibility of B's doing so-and-so does not show that the function of A is to do such-and-such. A's behaviour

may indeed make B's behaviour possible, but it need not be there for that purpose. Homeostatic mechanisms, both in artefacts (e.g. thermostats) and in organisms (e.g. the operation of sweat glands), exemplify such causal relations, but that is not the reason they are deemed purposive. Thermostats have a function in virtue of design that aims at a good (the purpose for the sake of which the artefact exists), and homeostatic mechanisms in nature have a purpose because of their role in the maintenance of some feature that is necessary for the life of, is protective of, or beneficial for, the organism of the type in question. The concept of function here, in the sense that concerns us, gets a grip only where the benefit of a being is appropriately involved. (Hacker, 2007, p. 167)

In a hackneyed example, hearts are necessary for the distribution of oxygen and nutrients. To be sure, this consequence may be legitimately appealed to account for the heart's existence. This consequence is the result of teleology, but it is not in itself a teleological explanation. It is only to the extent that the distribution of oxygen and nutrients has almost instant benefits that this consequence is truly functional. Hearts are also necessary for the production of heart attacks, but this is not their function. Therefore, the claim that the existence of a trait can be explained by appealing to the effect of some activity may simply be considered a conditional, rather than teleological, explanation.

For different reasons, then, while the original autopoietic model and successive attempts to update it (Thompson, 2007; Varela, 1997; Weber & Varela, 2002), as well as to expand it into adaptive (Di Paolo, 2005) or autonomous perspectives (Moreno & Mossio, 2015), have enabled a deeper understanding of teleology, it could be argued that the notion may require further analysis to better reflect the complexity of life processes. So, is there a variant of self-organisation theory that portrays teleology as an emergent biological principle that clearly describes the benefit of an individual beneficiary, while retaining the most valuable ideas of the Kantian and autopoietic tradition?

#### 4. Deacon's teleodynamics

In this final section, I will briefly argue that, despite the difficulties of theory discussed above, teleology can be accommodated in a variant of self-organisational theory that, so to speak, uses self-organisation against itself to create new and unprecedented forms.

To this end, this section will discuss two related issues: (a) the emergence of teleodynamic processes from self-organising dynamics. To provide a glimpse of how this might be possible, the theory has formulated an origin-of-life model based on autocatalysis; (b) the theoretical underpinnings of teleodynamic theory.

While most of this section focuses on (a), the most crucial difference between the OA and teleodynamics rests on (b) – that is, the account of how teleology distinctively operates, as opposed to how it emerged.

Of all the theories inspired by autopoiesis and by the theory of complex dynamical systems, Deacon's (2012) teleodynamics (Deacon & Cashman, 2013) is the only one that describes teleology as a natural cause and is distinctively rooted in far-from-equilibrium thermodynamics. 'Teleodynamics' is a newly coined term that designates a higher order relationship that can emerge in interactions among self-organising processes, that is, non-linear dynamical regularities that are persistently driven far-from-equilibrium. Taken in isolation, and individually considered, none of these self-organising processes are end-directed. Even less can it be said that they make an actual individual. Deacon argues that on the basis of this emergence, there is a phase transition from self-organisation to teleodynamics. This transition is discontinuous in the sense that it is characterised by a reversal of preceding dispositions from processes that become more orderly, or more constrained, by maximising the throughput of matter and energy, to processes that utilise but limit this throughput to prevent the loss of order. Teleodynamic processes thereby reduce or stabilise entropy production by regulating the possible dynamics of the system, and the ways this energy can degrade.

Teleodynamics depends on and is sustained by self-organising, far-from-equilibrium dynamics, but takes advantage of differences between self-organising processes to compensate for their intrinsic limitations. Because of this, teleodynamics inverts the basic logic of autopoietic approaches and is qualitatively different from them. Rather than focusing on the closure of constraints and the component generation processes, teleodynamics focuses on the prevention of system degradation, that is, on the ubiquitous risk of succumbing to the second law of thermodynamics, as I will now elaborate.

Following the pioneering work of Prigogine in the late 1970s, many theorists have understood living systems as dissipative structures. When pushed far from equilibrium, a system that initiates non-linear interactions among its components exhibits a dissipative structure. These recursively amplify the efficiency with which the system degrades the forces driving it away from equilibrium. But for this reason, dissipative structures tend to deplete the very energetic or material gradients that drive them. As a result, they most efficiently destroy themselves. By increasing local constraints and regularity, that is, order, self-organising processes obey the principle of MEP globally. This principle asserts that a system organises itself to maximise the rate of entropy production. In other words, to maintain its

order inside and to reduce entropy, self-organising systems rapidly degrade available energy gradients until it exhausts them, and so, faster increase in global entropy.

The MEP principle does not merely refer to the way in which systems tend towards maximum entropy following the second law of thermodynamics. It refers rather to processes that persistently achieve the maximum rate of entropy production. When a system is being constantly driven away from equilibrium, asymmetries of energy and material continually build up. As a result, this produces an increase in the intensity of the tendency to rapidly dissipate this difference. This increased potential does work that organises this dissipation process in a way that minimises the pathways by which this can occur, thereby constraining its ‘flow’ and producing local regularities – that is, order.

How could life depend on dynamics that form in the process of most efficiently destroying their necessary boundary conditions? Teleodynamic theory answers this fundamental challenge by envisioning a way that two or more MEP self-organising processes can become linked in a way where each prevents the other(s) from reaching this end. These processes do not merely reciprocally facilitate one another, but also prevent each other from completely degrading the energetic and material gradients that they each depend on. In this respect, they are both exporting entropy to minimise it internally, while also regulating this process to prevent loss of this capacity. In this way, the component processes stop each other’s depletion of energy gradients and counter the MEP principle that characterises dissipative systems. By limiting the degrees of freedom of a system, the linked self-organising processes build new and entropy-resistant structures that provide better and long-term stability. The emergence of such structures is only possible because the boundary conditions of each self-organising process are produced by the other(s) and therefore come to be internal, in a system of mutually supportive and mutually inhibiting constraints.

To provide an account of how teleodynamics might have spontaneously emerged from self-organising tendencies and an empirically testable model, Deacon (2012) devised the so-called ‘autogen model’. Since the details of autogenesis have been elaborated elsewhere (Deacon & Cashman, 2013; Koutroufinis, 2013; Sherman, 2017), I will briefly describe it. The model is inspired by the possibility of a hypothetical non-parasitic – that is, self-replicating – virus, that is, an autogen. All known viruses are parasitic on the molecular machinery of other living cells to replicate their nucleic acid (RNA or DNA) content and capsid containment molecules. Viral capsids spontaneously self-assemble into containers for their nucleic acid genes because of the symmetry of the component capsid molecules. But if a set of reciprocally catalytic molecules – in which each produces another in the set such

that all get produced – is replaced for the typical nucleic acids, molecular replication is accomplished without nucleic acids. And if one side-product of this reciprocal catalysis produces a capsid forming molecule, capsid formation will tend to occur where reciprocal catalysis is most rapid. Consequently, it will tend to enclose the molecules of the autocatalytic set. If the containment breaks open in an environment where sufficient catalytic substrates are available, then, the whole molecular complex will be likely to reassemble, giving rise to fresh tokens of the self-enclosed autocatalytic set that, in effect, replicate the autogenic system.

The autogen model describes two or more self-organising or constraint-generating processes<sup>3</sup> that become mutually coupled in such a way that each both supports and impedes the other’s MEP tendency, managing to create a stable set of molecular processes. By keeping each other within a narrow range, the coupled self-organising processes manage to create a stable individuated molecular system. Although the process of impeding entropy production can terminate in an inert molecular structure, when external perturbation breaks the structure and this loses integrity, this reinitiates the codependent self-organising processes and results in self-repair or self-reconstitution – and even self-reproduction – if its molecular components are dissociated but still able to interact.<sup>4</sup> This is possible because the codependent constraints are distributed among its different components irrespective of the loss of integrity of the whole. This distribution enables maintaining a stable set of mutually codependent constraints over time despite general instability, loss and replacement of components in the autogenic system. Together, the preservation of the system of relations between codependent processes and the capacity to reproduce these constraints in different substrates also provides the minimal conditions for evolution.

More importantly, teleodynamics is a process with the ontological role of promoting its persistence with respect to the incessant risk of dissolution due to extrinsic perturbation. The process is marked by the emergence of dynamical tendencies that act towards something that is not present in far-from-equilibrium processes, in their components or in the sum total of their components. Why is this process end-directed? Because it is organised with respect to something other than those components, that is, the constrained integrity of the individuated whole – its beneficiary. Taken in isolation, each component process of teleodynamics does not do any end-directed work; they are just self-organising. Because there is no beneficiary in raw self-organising dynamics, the emergence of teleology, while in continuity with it, is the result of the reversal of preceding dispositions.

Because, at first glance, teleodynamics might appear to be as no more than a straightforward refinement of closure, or the unpredictable complexification of self-

organising tendencies, one might easily be led to assume that the theory provides no more than a refined type of self-organisation, or that its distinction with the OA is merely semantic. Yet, its understanding of the key phase transition in purely normative terms is a sufficiently radical change as to merit reconsideration of this assumption.

The phase transition from self-organisation to teleodynamic organisation exhibits the distinctive features of life. Specifically, this transition describes the rise of processes whose goal is to stabilise runaway tendencies and engage in self-repair, system replication and consequence-organised work. As argued, a self-organising system that is ultimately maintained by the development of far-from-equilibrium tendencies fails to make a causally closed individual. Here is the reason. While the MEP principle requires constant extrinsic perturbation to maintain the system, its fundamental disposition is self-elimination. In this respect, the most critical boundary conditions of this system are extrinsic, and its most distinguishing disposition is self-destruction. To understand individuated self as an intrinsic physical property, it must do self-maintenance work, opposing the overriding tendency of MEP processes to undermine self-organisation. Hence, the emergence of an individuated self requires a phase transition from self-organisational systems that merely stay far from equilibrium to systems that act with the specific purpose of preventing their dissolution over time. This is why whirlpools, tornados, candle flames, autocatalytic sets or Bénard cells cannot be considered teleological.

To be sure, the OA account describes a phase transition from self-organisation to self-determination in which boundary conditions come to be internalised. The theory also envisages a dependence relation among constraints when system integrity is at risk. Constraints are said to only be so at specific time scales in which they are open to thermodynamic flow. At other times, they seem to be the outcome of processes guided by other constraints (Moreno & Mossio, 2015, pp. 18–19) to allow for self-repair.

Despite this, we might ask the following: why runaway MEP tendencies do not overwhelm self-maintenance work? Why are dissolution tendencies successfully countered by contrary tendencies? The teleodynamic system self-regenerates, and ultimately succeeds in opposing MEP tendencies because it is teleologically organised to prevent its dissolution.

Different advocates of the OA have held different views concerning the ability to neatly distinguish self-organising from self-determining dynamics. For one thing, Mossio and Bich are ready to contemplate –if somewhat reluctantly –and even accept the possibility that dissipative structures constitute a non-biological case of self-determination (Mossio and Bich, 2017, p. 1109), thereby signaling the possibility that these structures be teleological. This exploratory move may hint at

the existence of internal ambiguities within the view of different representatives of the OA about the dividing line between self-organisation and self-determination.

In contrast, the phase transition described by teleodynamics moves beyond self-organisation in a more radical way: by utilising MEP, the system prevents it from undermining the constraints that it generates and promotes self-beneficial work. The key phase transition from self-organisation to self-determination that ushers a new causal regime is the work done for the sake of the individual or its greater benefit.

Recall the questions raised in the introduction concerning (a) individuation, (b) agency and (c) benefit. We are now in a position to look at how teleodynamics answers them.

What is (a) individuation? As argued, teleodynamics describes individuation as the interlocking of at least two self-organising processes that oppose each other's MEP tendency as well as a tendency for them to decouple. This creates a system that tends towards stability despite the ubiquitous threat of degradation. Note that the opposition of the two self-organising processes is consequence-organised, that is, end-directed towards generating certain features, the most central of which is a system's capacity to preserve its own constraints – but, as argued, this is not the only one. Critical to the theory of teleodynamics is the idea that a biological system does not acquire its identity simply because its internal constraints realise closure – and ultimately, autonomy or self-determination. Rather, constraints are organised, that is, constrained themselves by a higher order, more powerful constraint for the sake of achieving specific consequences.

What is (b) agency? This feature is the result of the system's ability to 'perpetuate the reciprocal maintenance of the constraints that maintain the organism' (Deacon, 2012, p. 480). According to the MEP principle, the regularity produced by a self-organising process is a disposition towards degradation – only constant perturbation maintains the regularity. In contrast, by limiting the degrees of freedom of biological processes, teleodynamic constraints enable only forms of work that preserve, amplify and propagate those same constraints. In this sense, it is possible to say that the system does not sit idle, but does work to preserve and protect its integrity, confronting damage and perturbation with contrary work. Doing so, living processes genuinely act on their own behalf.

Finally, (c) biological work requires a beneficiary, an individual with respect to which some tendencies and consequences are promoted, and other which threaten or undermine its integrity relegated. Because teleodynamic systems are organised to achieve specific outcomes that are different from its component processes, they act for the sake of a beneficiary. All teleodynamic work is done for the sake of it, so that the capacity of the system – its self – be preserved. In this way, the

teleodynamic system is normative: amid constant perturbation, there is something that the system must do to preserve its integrity.

How does teleodynamics answer the questions raised in Section 3 in dialogue with the OA? Let us remind ourselves of them. It was argued that (a) biological individuality may be better determined, (b) thermodynamic work must involve a beneficiary and (c) the critical difference between a conditional and a teleological explanation should be noted.

In the case of teleodynamics, (a) the theory defines individuality as the interlocking of at least two self-organising processes that oppose each other's MEP tendency as well as a tendency for them to decouple. They both produce each other's boundary conditions. When successful, this process gives rise to a new a second-order constraint that sits above the constraint system. This constraint was absent from the component processes and must be specifically distinguished from them. Its goal is to preserve system integrity by resisting disintegration of their codependent component processes (Deacon & Cashman, 2016, p. 418). This is how a system of constraints come to be a biological individual. This introduces a crucial difference between systems of components that are merely circularly supportive of one another – as are the components of an autocatalytic set or of any other circularly organised production process – and teleodynamic systems. The second-order constraint is not a mere aggregate of interlocking constraints; it is also a full individual that belongs to a biological class. The critical reason why biological systems are identifiable as a unified class is the existence of this emergent, second-order constraint. (b) Teleodynamics describes a mode of physical causality that is inspired by Aristotle's theory. The teleodynamic account undermines the distinction between the functional and the operational level as delineated in Maturana and Varela's account, to claim that teleology is indeed a distinctive natural cause. In this way, teleology is deemed more than just a way to justify functional claims or the conditions for the continuity of the system, even if they both reflect part of the effects of what it does. Rather than describing the circularity of self-enabling processes, teleodynamics describes a radical phase transition that opens up the possibility of normative attributions of value and of the system welfare. By stressing this dimension, teleodynamics does not explain the attribution of function in the way of standard teleosemantic theories. By moving the discussion to the dynamics of nature itself, the theory aims to ground a truly teleonaturalistic account where the emphasis is placed on how nature behaves to ground normative descriptions of value. Finally, (c) teleodynamics does not ignore the fundamental relationship between conditional and teleological explanation. By treating the good of a beneficiary as a core defining attribute, teleology is not the resulting end of some

antecedent conditions. Processes are oriented to their telos in an unmistakably normative way to ensure that the system arrives at specific outcomes (of course, these outcomes are subject to what Di Paolo's calls a precarious autonomy. Processes are always beset by contingencies that may frustrate their telos).

In the last analysis, even though theories in the tradition of self-organisation describe constraint regularisation, closure, and self-maintenance processes, what sets teleodynamics apart from these is that it places teleology back on centre stage. The theory of teleodynamics argues that teleology is not merely a symbolic description or functional ascription assigned to a physical process from an extrinsic perspective. Nor can it be viewed as the net result of increased self-organisational complexity. And neither is it the end of self-organisation and far-from-equilibrium dynamics, since this would identify it with the MEP principle, and in the last analysis, with the processes that drive its own elimination. The telos of a living process is the consequence-organised work directed to maintain integrity in the face of uncertain possible future conditions, and the capacity to increase the probability of beneficial consequences despite the ubiquitous tendency towards degradation.

I argued that teleodynamic processes emerge from and thereby become the defining properties of biological self. So, it might be objected, does teleodynamics not assume benefit and normativity, leaving the account ungrounded and circular? In my view, the claim that teleology needs to be grounded on an account that lacks these features misses the point that they must nevertheless be accounted for in terms of the physical self-repair and self-regulation (i.e. constraint on runaway entropy production), which are the hallmarks of teleodynamic processes. Both self-repair and self-regulation assume that work must be constantly performed to minimise deviation from a target state. Teleodynamics defines this target state as the physical configuration in which these dispositions are maintained. So, defining the self as the instantiation of this circular dispositional state provides an individuated beneficiary and avoids circularity in the definition of normative relationships. Certainly, autopoiesis and the OA tacitly assume a preferred physical state of organisation, but its physical individuation and relation to these processes lacks an intrinsic disposition to achieve this target state, thereby leaving normative relationships undefined.

Despite this, teleodynamics is not immune to every possible charge. For one thing, Deacon's (2012) original autogenic model (Deacon & Cashman, 2013) did not recognise the problem of biological individuality as central to the theory. The dependence relation between teleology and biological individuality only became so in subsequent work (Deacon & Cashman, 2016). The details that bound up key features of the teleodynamic system such as constraint preservation throughout

change, the need for a beneficiary of teleodynamic work and the emergence of a higher order constraint are only implicit in *Incomplete Nature*, and demand some integrated analysis. In this context, while emphasis was made on the idea that a biological individual is the product of the interlocking of at least two self-organising processes, implicit in Deacon's work is the idea that a teleodynamic system is not the mere product of this interlocking, but the work of a beneficiary that acts to maintain its individuation. This idea seems critical to understand how the processes that spawn the rise of the autogen are not a complexification of self-organising dynamics, but a fundamental dynamical change.

## 5. Conclusion

Kant's original insight that organisms are intrinsically purposive self-organising systems was enormously influential. It anticipated and informed the modern understanding of self-organising processes and the development of autopoietic theories of living dynamics. But it also left important questions unanswered about the meaning of teleology that were inherited by later system-based theories. He believed self-organisation to be the distinctive feature of living forms, and was undoubtedly right in this, but he left open the question of whether the distinctive causality of life could be understood in physical terms or only descriptively. The contemporary developments of far-from-equilibrium thermodynamics that have led to an understanding of self-organising systems rekindled the belief that autonomy theories, which are open to far-from-equilibrium thermodynamics without being mechanistic, might at last give us full characterisation of natural purposes. Varela, Weber, Thompson, Di Paolo, Moreno, Mossio and many others have adopted strategies based on these premises. Based on Deacon's theory of teleodynamics, however, this article has argued that by inverting the focus from self-making to self-dissolution prevention, teleodynamics offers the glimpse of a path forward that restores Aristotle's final causality, accommodates intrinsic normativity and provides a promising Newton of the blade of grass.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The work done by the author was supported by the Mobility Program of the Spanish Ministry of Science and Innovation for Senior Academics for a stay at the University of California, Berkeley in 2019.

### ORCID iD

Miguel García-Valdecasas  <https://orcid.org/0000-0001-6271-5181>

### Notes

1. I am aware that the theory of biological autonomy assigns a more decisive role to the interaction of autopoietic systems with the environment, conceives autopoietic closure in terms of constraints (Meincke, 2018; Moreno & Mossio, 2015) and critically distinguishes the activity of self-organising systems like whirlpools or tornados, from that of autonomous systems. Its main features will be discussed in the next sections.
2. Deacon calls the general class of processes that include self-organisation 'morphodynamic' processes.
3. The two classes of self-organising molecular processes are (a) reciprocal catalysis – as described in the previous sections – and (b) molecular self-assembly – such as involved in the formation of cell membranes, virus capsids and crystallisation more generally. An autogenic system is created when side-products of the reciprocal catalysis tend to spontaneously self-assemble into a structure that can enclose the relevant collection of catalysts, thereby providing a unit structure with all the necessary codependent components required to reinitiate autogen reconstitution if damaged (for more details, see Deacon, 2012, ch. 10).
4. Reproduction in this sense is just a variant form of self-reconstitution. This is possible because the codependent constraints are distributed among its different components irrespective of the loss of integrity of the whole.

### References

- Allen, C., & Neal, J. (2020). Teleological notions in biology. In: E. N. Zalta (ed.), *The stanford encyclopedia of philosophy*. <https://plato.stanford.edu/archives/spr2020/entries/teleology-biology/>.
- Barandiaran, X. E., Di Paolo, E., & Rohde, M. (2009). Defining agency: Individuality, normativity, asymmetry, and spatio-temporality in action. *Adaptive Behavior*, 17(5), 367–386.
- Bedau, M. (1991). Can biological teleology be 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.
- Bich, L., & Damiano, L. (2008). Order in the nothing: Autopoiesis and the organizational characterization of the living. *Electronic Journal of Theoretical Physics*, 2008, 339–369.
- Christensen, W. D., & Bickhard, M. H. (2002). The process dynamics of normative function. *The Monist*, 85(1), 3–28.
- Deacon, T. (2012). *Incomplete Nature. How mind emerged from matter*. W.W. Norton.
- Deacon, T., & Cashman, T. (2013). Teleology versus mechanism in biology: Beyond self-organisation. In: A. C. Scarfe, & B. G. Henning (Eds.), *Beyond mechanism: Putting life back into biology* (pp. 287–308). Rowman & Littlefield.
- Deacon, T., & Cashman, T. (2016). Steps to a metaphysics of incompleteness. *Theology and Science*, 14(4), 401–429.
- Di Paolo, E. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, 4, 429–432.
- Di Paolo, E. (2009). Extended life. *Topoi*, 28, 9–21.

- Froese, T., & Stewart, J. (2010). Life after ashby. Ultrastability and the autopoietic foundations of biological autonomy. *Cybernetics and Human Knowing*, 17(4), 7–50.
- Garson, J. (2019). *What biological functions are and why they matter*. Cambridge University Press.
- Gilson, E. (1984). *From Aristotle to Darwin and back again*. University of Notre Dame Press.
- Hacker, P. M. S. (2007). *Human nature: The categorial framework*. Blackwell.
- Jacobs, J. (1986). Teleology and reduction in biology. *Biology and Philosophy*, 1, 389–399.
- Jonas, H. (1966). *The phenomenon of life*. Harper & Row.
- Kant, I. (1987). *Critique of judgment* (W. S., Pluhar, Trans.). Hackett. (Original work published 1790)
- Kenny, A. (1988). Cosmological explanation and understanding. In L. Hertzberg, & J. Pietarinen (Eds.), *Perspectives on human conduct* (pp. 72–87). Brill.
- Koutroufinis, S. (2013). Teleodynamics: A neo-naturalistic conception of organismic teleology. In A. C. Scarfe, & B. G. Henning (Eds.), *Beyond mechanism: Putting life back into biology* (pp. 309–342). Rowman & Littlefield.
- MacLaughlin, P. (2001). *What functions explain*. Cambridge University Press.
- Maturana, H., & Varela, F. J. (1973). *De máquinas y seres vivos: una teoría sobre la organización biológica*. Editorial Universitaria.
- Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living*. *Boston studies in the philosophy of science* (Vol. 42). D. Reidel.
- Meincke, A. S. (2018). Autopoiesis, biological autonomy and the process view of life. *European Journal for Philosophy of Science*, 9, Article 5.
- Moreno, A., & Mossio, M. (2015). *Biological autonomy: A philosophical and theoretical enquiry*. Springer.
- Mossio, M., & Bich, L. (2017). What makes biological organization teleological? *Synthese*, 194(4), 1089–1114.
- Nguyen, A. (2021). A functional naturalism. *Synthese*, 198, 295–313. <https://doi.org/10.1007/s11229-018-02002-x>
- Ruiz-Mirazo, K., & Moreno, A. (2012). Autonomy in evolution: From minimal to complex life. *Synthese*, 188(1), 21–52.
- Sherman, J. (2017). *Neither ghost nor machine*. Columbia University Press.
- Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Harvard University Press.
- Varela, F. J. (1979). *Principles of biological autonomy*. North Holland: Elsevier.
- Varela, F. J. (1991). Organism: A meshwork of selfless selves. In A. I. Tauber (Ed.), *Organisms and the origins of self* (pp. 79–109). Kluwer.
- Varela, F. J. (1997). Patterns of life: Intertwining identity and cognition. *Brain and Cognition*, 34, 72–87.
- Varela, F. J., Maturana, H., & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems*, 5(4), 187–196. [https://doi.org/10.1016/0303-2647\(74\)90031-8](https://doi.org/10.1016/0303-2647(74)90031-8)
- Von Uëxkull, J. (1982). The theory of meaning. *Semiotica*, 42, 25–82.
- Walsh, D. M. (2015). *Organisms, agency and evolution*. Cambridge University Press.
- Weber, A., & Varela, F. J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences*, 1, 97–125.

### About the Author



**Miguel García-Valdecasas** is an Associate Professor in the Department of Philosophy at the University of Navarra (Spain) and a leading member of the project Mind-Brain: Biology and Subjectivity in Philosophy and Neuroscience at the Institute for Culture and Society (ICS), University of Navarra. Drawing on Aristotle's philosophy of mind and action, his research has largely focused on the epistemological and philosophical challenges of neuroscience, and particularly on the role of teleology in the problem of mental causation.