

Modeling Autopoiesis and Cognition with Reaction Networks

Heylighen, Francis; Busseniers, Evo

Published in:
BioSystems

DOI:
[10.1016/j.biosystems.2023.104937](https://doi.org/10.1016/j.biosystems.2023.104937)

Publication date:
2023

License:
CC BY-NC-ND

Document Version:
Accepted author manuscript

[Link to publication](#)

Citation for published version (APA):
Heylighen, F., & Busseniers, E. (2023). Modeling Autopoiesis and Cognition with Reaction Networks. *BioSystems*, 230, [104937]. <https://doi.org/10.1016/j.biosystems.2023.104937>

Copyright

No part of this publication may be reproduced or transmitted in any form, without the prior written permission of the author(s) or other rights holders to whom publication rights have been transferred, unless permitted by a license attached to the publication (a Creative Commons license or other), or unless exceptions to copyright law apply.

Take down policy

If you believe that this document infringes your copyright or other rights, please contact openaccess@vub.be, with details of the nature of the infringement. We will investigate the claim and if justified, we will take the appropriate steps.

Modeling Autopoiesis and Cognition with Reaction Networks

Francis Heylighen & Evo Bussieniers
Center Leo Apostel
Vrije Universiteit Brussel

Abstract: Maturana and Varela defined an autopoietic system as a self-regenerating network of processes. We reinterpret and elaborate this conception starting from a process ontology and its formalization in terms of reaction networks and chemical organization theory. An autopoietic organization can be modelled as a network of “molecules” (components) undergoing reactions, which is (operationally) *closed* and *self-maintaining*. Such organizations, being attractors of a dynamic system, tend to self-organize—thus providing a model for the origin of life. However, in order to survive in a variable environment, they must also be resilient, i.e. able to compensate perturbations. According to the “good regulator theorem” this requires some form of *cognition*, i.e. knowing which action to perform for which perturbation. Such cognition becomes more effective as it learns to anticipate perturbations by discovering invariant patterns in its interactions with the environment. Nevertheless, the resulting predictive model remains a subjective construction. Such implicit model cannot be interpreted as an objective representation of external reality, because the autopoietic system does not have direct access to that reality, and there is in general no isomorphism between internal and external processes.

Keywords: autopoietic systems; resilience; chemical organization theory; origin of life; reaction networks; cybernetics; self-organization; cognition; epistemology; process ontology.

Introduction

Half a century ago, the Chilean biologists Humberto Maturana and Francisco Varela introduced a revolutionary perspective on what it means to be a living system: *autopoiesis* (Maturana and Varela, 1980; Varela et al., 1974). For them, the essence of life is that it produces (*poiesis*) itself (*auto*). A living system is autonomous: it determines its own structures and processes, thus ensuring that its organization would maintain independently of what happens in its environment. Such active self-

determination is to be contrasted with the passivity of non-living systems, such as a heap of sand or a drop of water, whose structure and evolution is purely the result of external forces, such as wind, rain, or gravity.

Autopoiesis entails a number of associated aspects: an unambiguous distinction or separation between the autonomous system (self) and the rest of the world, a circular or self-referential form of organization, and the ability of the system to compensate for external events that would otherwise disturb its organization. Unlike a heap of sand, a living system will actively counteract, neutralize, or evade potentially damaging external influences, thus safeguarding its continued existence under variable circumstances.

For Maturana and Varela (1980), such compensation of perturbations implies *cognition*: the organism must be able to sense, recognize and adequately react to perturbations by performing the right counteractions. Vice-versa, cognition, in the sense of intelligently responding to perceived phenomena, only makes sense for autonomous, living systems. Thus, for them, *autopoiesis* and *cognition* are two sides of the same coin: the one cannot exist without the other. Later authors (e.g. Bitbol and Luisi, 2004; Bourguin and Stewart, 2004) have tended to separate these aspects, conceiving of primitive forms of autopoiesis that do not exhibit cognition (e.g. the self-assembly of cellular compartments), or of forms of cognition that do not require autopoiesis (e.g. AI programs). In the following, we will argue for a continuum between the simplest, more rigid forms of self-production and the more intelligent forms that appear to exhibit cognition. Thus, for us, cognition is the natural continuation of autopoiesis, albeit that it only comes to the fore as the autopoietic system evolves to become increasingly adaptive, or, as we will call it, *resilient*.

Next to clarifying the link between autopoiesis and cognition, the present paper will focus on an aspect of the theory of autopoiesis that has as yet received less attention, which is that it presupposes a *process ontology* (Meincke, 2018). The standard ontology of science, which goes back to Aristotle, reduces all phenomena to *objects*: static, material components, such as particles, molecules or stones, that may be moved or rearranged by external forces, but that retain their essential properties, independently of the processes in which they participate (Heylighen, 2023; Walsh, 2018). For example, traditional, Neo-Darwinist accounts of living systems tend to reduce them to assemblies of genes, i.e. invariant DNA molecules. Let us contrast these reductionist and materialist assumptions with the often-cited original definition of autopoiesis (Maturana and Varela, 1980; Varela, 1979; Bich and Etzeberria, 2013):

An autopoietic system is organized (defined as a unity) as a network of processes of production (transformation and destruction) of components that produces the components which:

(1) through their interactions and transformations regenerate and realize the network of processes (relations) that produced them; and

(2) constitute it (the machine) as a concrete unity in the space in which they exist by specifying the topological domain of its realization as such a network

To clarify this notoriously complex statement, we will for the time being ignore condition (2), and unpack the first part of the definition into simpler propositions:

- An autopoietic system is organized as a network of processes.
- These processes produce, consume and transform the components of the system.
- These components interact so as to regenerate the network of processes that produced them.

Let us further reduce this description to its essence: *the system's processes produce its components, which in turn generate its processes*. Given that they are being consumed, produced and transformed, the components here should not be seen as static objects, but as raw materials or intermediate stages within an on-going processing cycle. The only thing that maintains is the network of processes, which is continuously regenerating itself. This is the essence of “self-production”: while a process by definition transforms something into something else, the eventual output of the sequence of processes must again include the input from which that sequence started, so that the overall process continues running.

Thus, autopoiesis entails a cycle or *loop*, famously depicted as the Ouroboros, the snake that eats its own tail. This has inspired Varela and other autopoiesis theorists (Varela, 1979; Reichel, 2011) to investigate the seemingly paradoxical notion of self-reference—such as the sentence that says of itself that it is false. A perhaps more practical illustration of such circularity is the feedback loop that supports the system's autonomy: by producing its own input or initial conditions, an autopoietic system makes itself largely independent from the environment (Heylighen, 2022)—which would otherwise produce the inputs or initial conditions that determine the course of the dynamical process.

Of course, thermodynamics requires autopoietic systems to be energetically open: they need an input of low-entropy resources (“food”) to sustain the on-going process, while producing an output of high-entropy “waste”, which they dissipate in their environment. But this input does not determine the processes that take place inside the autopoietic system: the system is *operationally* (or *organizationally*) *closed*. That means that these processes or operations do not change the essential organization of the system: the network of cause-effect relationships in which they take part remains the same throughout the operation of the system.

What does change, however, is what Maturana and Varela call the “structure” of the system. This includes the particular arrangement of its physical components, such as cells, vessels, or leaves. Components may be broken down, transformed or newly produced, so that the material composition of the system changes. This may be necessary in order to cope with changing conditions in the environment. Yet, the essence is that the process of autopoiesis is not interrupted, and therefore that the autopoietic system maintains its unity or identity within a universe of change.

It is this continuity of identity across on-going change that we here wish to focus on, and this while starting from processes as fundamental elements. As

explained further, these processes can be formally modelled as reactions. We will show under which conditions networks of reactions can be understood as autopoietic. That will allow us to formalize and model the core concepts of the theory of autopoiesis.

It must be noted that previous attempts at formalizing autopoiesis have mostly focused on its topological aspect, i.e. the formation of a boundary surrounding an autopoietic system (McMullin and Varela, 1997; Varela et al., 1974). A frequently used formalism for describing the formation of such spatial structures is a two-dimensional cellular automaton—such as the “game of life” (Beer, 2020). Our approach will ignore this spatial aspect, while focusing on less easy to visualize features, such as organizational closure, self-maintenance, and cognition, arguing that with the proper ontological assumptions and formal representation these features can be explained simply and transparently.

A fundamental benefit of our approach is that it will allow us to model the *origin* or *emergence* of autopoiesis: how can initially independent processes become organized into a unified, self-producing network? This origin of autopoiesis can also be seen as the origin of life (Bourgine and Stewart, 2004), autonomy, agency or goal-directedness (Heylighen, 2022; Heylighen et al., 2022). As yet, autopoiesis-inspired origin-of-life scenarios have mainly focused on the formation of an enclosing boundary, i.e. a primitive cell wall. Such a scenario has even been demonstrated in the laboratory, showing the spontaneous formation of “vesicles” that can grow and even multiply (Luisi, 2003, 1994).

However, the true wonder of life, for us, lies in its self-maintaining organization and its cognitive ability to adequately counteract environmental perturbations, thus safeguarding its organization. It is this ability that turns the autopoietic system into an autonomous agent, which is not dependent on its environmental conditions, the way a passive object is, but able to actively intervene in these conditions. That allows it to work towards situations that promote its further survival, while evading situations that endanger it. It is this *origin of goal-directed action* (Heylighen, 2022) that we will now try to explain by means of the reaction network formalism. Thus, this paper builds further on a previous study (Heylighen et al., 2022) by clarifying the function of cognition in autopoiesis, while providing a process-based interpretation of the non-representationalist epistemology that Maturana and Varela derived from these ideas (Maturana, 1995; Maturana and Varela, 1980)

Reactions as elementary processes

A process ontology assumes that the most fundamental building blocks of reality are *processes* (Bickhard, 2011a; Rescher, 2000; Seibt, 2022), rather than static objects, particles or substances. The standard, object-based ontology that has dominated Western science and philosophy since Aristotle (Walsh, 2018) is materialistic: it

assumes that objects consist of an inert *substance*, which we call matter, and that this matter ultimately consist of elementary components that we call *particles*. That allowed a reductionist strategy, in which complex phenomena, such as cells were analyzed into smaller and simpler components, such as molecules, thus facilitating modeling and understanding. Older process philosophies, such as the ones of Whitehead (1978), Teilhard de Chardin (1959) and Bergson (1911), did not systematically use this strategy, but were focusing on grand, encompassing processes with many interconnected aspects, such as evolution or ontogenetical development. That made modeling difficult. Therefore, these philosophies seem to have had little impact on science. For modeling, an analytic strategy remains useful, especially when trying to understand the origin of complex systems from simpler components. The approach we will use therefore starts from *elementary* processes (Finkelstein, 1973), which we will call “reactions”. These play the role of the elements or building blocks that constitute more complex processes—just as particles do in an object ontology.

A rationale for the primacy of elementary reactions over elementary particles can be found in quantum field theory (Bickhard, 2011b; Hättich, 2004; Kuhlmann, 2000), which is the most fundamental (and most accurate) theory we have in physics. Here, “particles” are conceived as merely temporary excitations of a field that describes potential interactions. Particles can be emitted or absorbed, react with other particles in order to produce new particles, or sometimes even emerge from a “quantum fluctuation of the vacuum”, i.e. out of nothing. The only thing that really counts in quantum theory is how a particle reacts with another particle or with a larger-scale system (such as an observation apparatus), because it are these interactions that ultimately allow us to observe the particle’s behavior or properties. Particles that do not interact simply cannot be observed. Therefore, according to quantum mechanics they do not have any determinate properties (and we may as well assume that they do not exist).

For example, seeing an object, such as a flower, assumes that photons absorbed from the surrounding light and selectively reemitted by the pigment molecules in the flower are absorbed again by the cells in our retina. There they trigger electrical excitations (burst of electrons) that travel from neuron to neuron across our brain, while releasing and reabsorbing neurotransmitter molecules within the synaptic clefts between neurons. Thus, our conception of the flower as a static object just “being there” outside ourselves is actually the outcome of a very complex neural process that searches for invariants within a whirlwind of circulating processes of propagating activation, whose origin lies partly in the ever-changing stimuli from the surrounding environment, partly in the never-ending activity of the nervous system.

Let us then try to reduce this complexity by conceiving of the simplest possible process. This could be the appearance of some distinguishable condition *a* (e.g. light). We will represent this appearance, production, or generation of *a* by the following reaction:

$\rightarrow a$

We can similarly represent the disappearance, consumption or absorption of a (e.g. onset of darkness) as:

$a \rightarrow$

Such processes on their own are not very useful yet, because we want to understand how conditions or processes interact, thus forming networks that may exhibit certain forms of invariant organization. The next simplest process is the transformation of some condition a into a different condition b :

$a \rightarrow b$

One way to interpret this is as a causation, with a as cause and b as effect. Another interpretation is what has been called a “condition-action rule” in Artificial Intelligence models of agents (Holland et al., 1989; Russell and Norvig, 2009). Whenever the condition a holds, some action is performed that changes this to the new condition b . That action may occur spontaneously, like when some change happens in the circumstances (e.g. rain cloud \rightarrow rain), or, as we will discuss later, it may be initiated by some “agent”.

This simplest level of formalism is already sufficient to define a minimal autopoietic network of reactions:

$\rightarrow a$
 $a \rightarrow b$
 $b \rightarrow a$
 $b \rightarrow$

Here, component a produces component b , while b again produces a . This is the simplest form of a cycle, in which a and b produce each other. To take into account thermodynamics, we must also include dissipation, here represented as the loss or disappearance of some of the b . To compensate for that loss, there must be some (external) input (generation) of a (see Fig. 1). Note that at this abstract level of modeling, the formalism does not specify where some condition, such as a , comes from; it merely notes that a is somehow entered into the system being modeled. In practice, for thermodynamic systems, there will need to be some external source supplying the “food” a that keeps the metabolic cycle running. On the other hand, the formalism can also be used to describe purely informational systems, in which there is no conservation law of matter or energy, in the sense that “ideas” or “insights” can appear or disappear without necessarily being processed into something else.

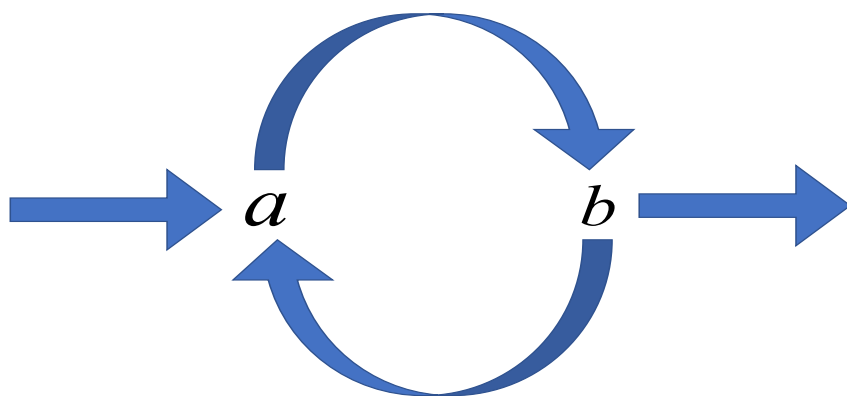
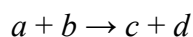


Figure 1: a minimal autopoietic network of reactions, in which *a* is processed into *b* and *b* back into *a*, while some of the *b* leaves the system, and some of the *a* enters from outside the system

This example is of course trivial, and as such not very illuminating in terms of understanding living systems. We could make it more complex by adding intermediate stages, such as: $a \rightarrow b$, $b \rightarrow c$, $c \rightarrow d$, $d \rightarrow a$, but this cycle remains a simple series of repeating intermediate conditions.

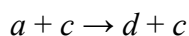
The formalism becomes more interesting when we note that conditions can have internal structure. That means that we can decompose complex conditions into conjunctions of simpler conditions: *a* and *b* together produce *c* and *d*. Using the conventional notation for reactions in physics and chemistry, we will write such conjunctions with a “+” sign, e.g.:



The simplest interpretation is that two or more specific conditions (here *a* and *b*) are jointly present, and that this combination triggers a process or reaction that transforms these conditions into a new combination of conditions (here *c* and *d*). These conditions could represent the presence of particles, molecules, organisms, physical structures, linguistic expressions, or social situations, or simply any kind of distinguishable conditions. Our process ontology does not care about what kind of “thing” it is that undergoes change, but only about what precisely the change consists of—i.e. in what way the resulting condition is distinct from the initial condition.

When investigating the origin of life, the most straightforward interpretation is that the conditions *a*, *b*, *c*, ... represent the presence of molecules that undergo chemical reactions. However, the reaction network formalism is not restricted to chemistry, and can be applied in the most diverse disciplines, including ecology, physics, computer science and sociology. As such, it provides a language for the transdisciplinary modeling of systems of the most diverse types (Heylighen, 2023; Heylighen et al., 2015; Veloz and Razeto-Barry, 2017).

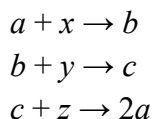
An important observation is that a condition c may be necessary for a reaction to occur, however, while not being affected by that same reaction. In that case, the condition c appears both before and after the \rightarrow sign that represents the process, as in:



In chemistry, c can be interpreted as a *catalyst*: a molecule that enables the reaction but that is not consumed by that reaction. Another possible interpretation is that c is the *agent* of the action that transforms a into d —the way an animal transforms food and oxygen into waste and CO₂, or a computer processes numerical data into diagrams. But c can also be a context, structure or background condition that must remain present for the reaction to take place. For example, falling is a process that moves a massive object from a higher position to a lower one, albeit under the condition that there is a gravitational field. In interstellar space, where gravity is negligible, the process of falling will not take place. Another example of a necessary condition for a process to occur is a connecting road, canal or tube that enables the transport of some substance from one place to another. The autopoiesis-inspired theory of biological organization developed by Alvaro Moreno, Matteo Mossio and colleagues refers to all such process-guiding structures and conditions as “constraints”, and characterizes living systems as being able to produce their own constraints (Montévil and Mossio, 2015; Mossio and Moreno, 2010)

In the remainder, we will call such enabling but stable conditions “catalysts”. Yet, while catalysts remain invariant during the reactions they catalyze, there will in general be other reactions that do consume or produce them. Thus, while catalysts introduce a form of local stability of conditions in the formalism, the general assumption is still that everything is subject to change, and in particular that autopoietic systems produce their own catalysts.

A common model for the origin of life is an autocatalytic cycle, in which certain molecules *indirectly* catalyze their own production, so that their number increases. A simple example is the following:



Here, a , b , and c are component molecules of a cycle that multiplies all these components while consuming the “food” molecules x , y and z . This could again be viewed as a simple, albeit somewhat more realistic, model of an autopoietic system. The cycle it forms is still a simple, linear succession of steps or stages, producing consecutively a , b , c , and then again a .

The metabolic reactions in actual organisms, however, are much more complex, with different reactions involving distinct yet overlapping types of molecules that occur in parallel (simultaneously) and not just sequentially. While the

overall organization is autopoietic, we can no longer write it down or visualize it as a simple cycle. That is why we need a more general mathematical criterion for recognizing a network of reactions as autopoietic. An elegant way to do this was proposed by Peter Dittrich, thus founding an approach known as *Chemical Organization Theory* (COT), which we will now discuss.

Chemical Organization Theory

Chemical Organization Theory (Dittrich and Fenizio, 2007) is built on top of the formalism of reaction networks as we just sketched it. More precisely, such a network consists of a set of *molecules* or *species*: $M = \{a, b, c, \dots\}$, together with a set of *reactions*: $R = \{r_1, r_2, r_3, \dots\}$. A reaction is defined as a mapping from a subset of molecules onto another such set:

$$r_i \in R: \text{Input}(r_i) \rightarrow \text{Output}(r_i),$$

Here $\text{Input}(r_i)$ and $\text{Output}(r_i)$ are both subsets of M . (Note that in the original definition, these are more generally assumed to be multisets, so as to allow for elements to appear more than once, as in $c + z \rightarrow 2a$, as is typical for chemical reactions. However, since we do not need to discuss such cases, we will for simplicity only assume ordinary sets in the following).

The molecules of the input set are called the “reactants” of the reaction. Unless they are catalysts, they are “consumed” (i.e. removed) by the reaction, and processed or converted into the molecules of the output set. These form the reaction’s “products”.

Note that the network defined by such reactions between molecules is not a directed *graph* (one-to-one connections), but a directed *hypergraph* (many-to-many connections) (Flamm et al., 2015). That is what will allow us to represent self-producing networks more complex than cycles.

Like in the previous section, we will write the elements of the input and output sets as a list of items separated by the “+” operator:

$$\begin{aligned} r_i: a + b + \dots &\rightarrow f + g + \dots, \\ \text{where } \{a, b, \dots\} &= \text{Input}(r_i) \text{ and } \{f, g, \dots\} = \text{Output}(r_i). \end{aligned}$$

The equivalent of an autopoietic system in COT is called a (*chemical*) *organization*. To define such an organization, we start with a subset O of the full set of molecules M , together with the subset $R(O)$ of all those reactions in R that are triggered by the presence of molecules in O . That means that $R(O)$ contains all those reactions r_i for which the input set $\text{Input}(r_i)$ is a subset of O .

$\{O, R(O)\}$ is defined to be an *organization* if and only if it satisfies the following two conditions:

- 1) *Closure*: this means that the output of any reaction in $R(O)$ is a subset of O . In other words, the reactions working on the given set of molecules O will not produce any new molecules that were not already in that initial set.
- 2) *Self-maintenance*: this means that all the molecules consumed by some reaction in $R(O)$ are also produced by some other reaction in $R(O)$, and this in an amount sufficient so that their overall concentration does not go down to zero.

Together, the two conditions specify that while the reactions in $R(O)$ are transforming the molecules into other molecules, the set of molecules available and the set of reactions taking place remain invariant: nothing new is added, nothing present is removed. That is exactly how we characterized autopoiesis in our paraphrase of Maturana's and Varela's definition:

- An autopoietic system is organized as a network of processes (here formalized as a *reaction network*).
- These processes (here formalized as *reactions*) produce, consume and transform the components of the system (here formalized as *molecules*).
- These components (*molecules*) in turn interact (*react with each other*) so as to again generate the network of processes (*reactions*) that produced them.

We still need to explain the last defining feature of autopoiesis, which we ignored up to now: the self-generating network must define the system as a *unity*, i.e. a coherent whole that is clearly distinguished or separated from its surrounding environment. Condition (2) of Maturana's and Varela's full definition suggests—and most subsequent treatments assume—that this unity is realized as a concrete compartment in space (such as a cell), which is physically separated from its surroundings by a topological boundary (such as a membrane). The reaction network formalism as yet does not include any notions of space or topology—although these can be added in various ways (Peter et al., 2021).

However, the definition of a chemical organization through its requirement of closure already implements a more abstract, algebraic form of distinct unity (Heylighen, 1990). That is because it unambiguously distinguishes between the molecules O that are part of the organization and all the other molecules in the “universe of discourse” M . Since these latter molecules by definition are not being produced or consumed by the processes that constitute the organization, they can only belong to its “outside”: they are alien to the organization, and not under the control of its processes. A remaining issue that we will explore in the sections on cognition is whether the organization can in some way still “know” that these molecules exist.

One advantage of such an algebraic notion of closure over a more concrete topological one is that it can be used to describe self-producing systems that are not localized in physical space. A classic (albeit controversial) application is to social systems, which according at least to the analysis of Luhmann are autopoietic networks of communications (Luhmann, 1986), even though they lack physical boundaries.

Chemical Organization Theory can indeed be used to model such Luhmannian autopoietic systems (Dittrich and Winter, 2008).

The main import of the notion of topological boundary is in understanding the origin of the first living cells. Autocatalytic or RNA-world scenarios for the origin of life (Heylighen et al., 2022; Hordijk et al., 2018; Pressman et al., 2015) focus on either self-production or replication of chemical structures, albeit without specifying the space in which the corresponding reactions take place. However, one difficulty with such scenarios is that in the real world molecules tend to diffuse, with the result that concentrations may become too low to sustain either autocatalysis or replication. Enclosing the reacting molecules within a semi-permeable membrane can prevent that problem, by ensuring that the internal concentration remains high enough for all the necessary reactions to take place (Deacon et al., 2014; Heylighen et al., 2022).

One of the more concrete models for the origin of autopoiesis (Bourgine and Stewart, 2004) proposes an autocatalytic cycle (similar to the one we sketched earlier), but where one of the products is a molecule that forms an encapsulating sheet, thus combining a self-producing network of chemical reactions with the production of a boundary. This boundary in fact plays the role of a catalyst for the reactions taking place within it, simply by ensuring that the molecule concentrations remain high enough to enable on-going reactions. This illustrates that even without a model of topology, the reaction network formalism can model the *functional* role of a membrane, simply by using different reactions to specify the dynamics of the system in either presence or absence of the membrane.

If such an assumption of membranes playing a catalytic role would be realistic, then we might expect that a chemical organization would be easy to evolve in the presence of some membrane generating process, while failing to arise in its absence. But to understand such scenarios better, we first need to examine another one of the main strengths of Chemical Organization Theory, namely its ability to model the self-organization of autopoietic systems.

The origin of autopoiesis

A reaction network can be seen as a *dynamical system*, with a state determined by the concentration of the different types of molecules, and a dynamical rule determined by the reactions working on these molecules. If we know the rate of the reactions (i.e. how much they produce or consume per unit of time), then we can calculate how the state changes over time (for a recent application, see e.g. Veloz et al., 2022). Such a dynamical system has typically one or more attractors. The attractors characterized by the sustained presence/activity of certain molecules and reactions are all chemical organizations—as formally demonstrated by Peter and Dittrich (2011). That makes sense, because reaction networks that are not (organizationally) closed or not self-maintaining will by definition gain or lose some of their molecule species, in contradiction with the assumption that in the attractor the set of molecules present in

the system is invariant. (Note that this invariant set can include molecules entering the system through a constant input of the form $\rightarrow a$, or exiting through a constant output of the form $\rightarrow b$. Closure of the organization here does not preclude thermodynamic openness).

A dynamical system, when left to evolve on its own, will tend to spontaneously settle into one of its attractors. Ashby (1962) has interpreted this as a basic principle of *self-organization*: once in the attractor, the different components of the system have lost their ability to move freely (and thus leave the attractor). That implies that their activities have to some degree become dependent on each other, defining some form of collective organization that maintains itself (Ashby, 1962; Heylighen, 2001). When the dynamical system is a reaction network, the resulting attractor is a chemical organization (autopoietic system) (Heylighen et al., 2022; Peter and Dittrich, 2011).

The conditions for such a system to self-organize depend on the initial state (molecules present) and dynamics (applicable reactions) of the system. Clearly, not all combination of molecules and reactions will give rise to non-trivial, self-sustaining networks. Most are likely to just settle into an equilibrium in which nothing much happens. Still, by simulating the dynamics of networks generated from random reactions, our research group has found that organizations (autopoietic systems) do emerge in quite a number of different conditions. Simulations are still on-going in order to determine more precisely what these conditions are (Veloz et al., 2022). But the general principle is simple enough to make it plausible that autopoietic organizations can self-organize rather easily—at least on the abstract level of the reaction network formalism.

The basic algorithm for simulating such a self-organizing process is that you begin with an initial set of randomly selected molecules. These then start reacting with each other while producing additional molecules. These react in turn producing even further new molecules. You then recursively add all molecules generated in this way to the set, until no new ones can be produced anymore by the reactions working on the set. The network now has by definition become closed. You then recursively remove the molecules that are consumed, but not (or insufficiently) produced, until nothing more can be removed. The remaining network is now self-maintaining as well as closed. That means that it is a chemical organization.

Let us illustrate this process using an example of a reaction network that was randomly generated to include 10 reactions, each producing or consuming between 1 and 3 resources out of a 10-element molecule set (see Table 1). Note that such random generation does not take into account thermodynamic constraints, such as conservation of matter and energy. That is because the “molecules” do not necessarily stand for physical entities: they can also represent abstract conditions that do not obey conservation laws. Similarly, reactions may represent abstract inferences, such as: if d is the case, then a must be the case as well: $d \rightarrow a + d$. For simplicity, we here consider a qualitative simulation, which only takes into account the presence or absence of molecules (not their precise concentration) as defining the system’s state.

$r1:$	f	\rightarrow	$f + g$
$r2:$	$j + a$	\rightarrow	h
$r3:$	$e + i + g$	\rightarrow	e
$r4:$	$a + g$	\rightarrow	$a + g + b$
$r5:$	$i + h$	\rightarrow	$c + a$
$r6:$	a	\rightarrow	$e + g + i$
$r7:$	d	\rightarrow	$a + d$
$r8:$	$e + c$	\rightarrow	$e + a$
$r9:$	e	\rightarrow	$f + d$
$r10:$	e	\rightarrow	$i + a$

Table 1. A reaction network consisting of ten randomly generated reactions $R = \{r_1, r_2, \dots, r_{10}\}$ that produce or consume ten molecules $M = \{a, b, \dots, i, j\}$

The process starts from an initial state, for which we can take an arbitrary subset of the molecule set M , for example $\{d, h\}$. The only reaction applicable in this state is r_7 , which adds the resource a to the set, producing the extended set $\{a, d, h\}$. This new state enables r_6 , which adds e, g and i , thus producing the next state $\{a, d, e, g, h, i\}$. Note that while r_6 consumes a , this does not immediately remove all a from the system, because we assume that a reaction has a finite rate. The presence of a and g activates reaction r_4 , which adds b to the set of molecules, while r_5 adds c . Thus, we now have: $\{a, b, c, d, e, g, h, i\}$. The presence of e activates r_9 which additionally produces f , resulting in $\{a, b, c, d, e, f, g, h, i\}$. The other enabled reactions (e.g. r_{10} or r_3) merely add molecules that are already there. The set has now become closed: no further molecules can be added by applying any of the reactions.

In this set, most of the elements are produced by some reaction working on other molecules in the set (e.g. the a consumed by r_6 is produced anew by r_7 and by r_{10}). However, h can only be produced by r_2 , which requires the molecule j that is not in the set. On the other hand, h is being consumed by r_5 . Therefore, h will eventually be eliminated from the set. Since c requires h for its production via r_5 , while being consumed by r_8 , it too will be eliminated. This leaves us with $\{a, b, d, e, f, g, i\} = A_1$. This 7-element set is closed and self-maintaining and therefore an organization. It is the attractor reached by the reaction dynamics starting from the initial state $\{d, h\}$.

Let us now start from the initial state $\{f\}$. Only one reaction is applicable, r_1 , producing the new set $\{f, g\} = A_2$. This set is closed, because no further reactions can be applied to it, and self-maintaining, because r_1 continuously reproduces it. Therefore, it is an organization (albeit a rather trivial one), and an attractor of the dynamics.

Resilience

The example of Table 1 can also be used to illustrate another important concept in the evolution of autopoietic systems: *resilience* (Heylighen et al., 2022; Veloz et al., 2022). For an autopoietic system to survive in a real, natural environment, it should be able to cope with changes in that environment. Normally, external conditions change,

independently of the internal dynamics of the autopoietic organization. That means that new molecules may be introduced to the overall reaction network, or that molecules present may be removed.

This would perturb the functioning of the autopoietic system if the molecules removed were being used by the autopoietic process, or if the new molecules would react with some of the molecules in that process. The first type of perturbation can be interpreted as a lack of “food” (resources necessary for self-maintenance). The second type can be interpreted as the introduction of “toxins” that interfere with the self-maintaining “metabolism” (Heylighen et al., 2022). On the other hand, if the molecules added or removed do not react with any of the molecules in the organization O , then the change will not affect the autopoietic process.

Thus, only certain changes in the overall reaction network will act as perturbations to the autopoietic system. For example, for an oxygen-breathing organism, such as the human body, lack of oxygen in the air is a potentially lethal perturbation, because oxygen plays a vital role in maintaining the metabolism. A somewhat more slowly working perturbation is the presence of carbon monoxide, CO, in the air, because CO reacts with the hemoglobin in the blood in such a way that this hemoglobin can no longer perform its function of transporting oxygen. On the other hand, if the nitrogen in the air would be replaced by another non-reactive gas, such as argon, then this would not have any effect on autopoiesis, because nitrogen does not react with any molecules used by the human metabolism.

For the autopoietic system to survive, it should be able to cope with such perturbations, i.e. adapt its internal processes to the changes in the concentration of available molecules in such a way that the overall process of self-maintenance is not interrupted. In the theory of autopoiesis and the related theory of cybernetics, this is called *compensation of perturbations*. The perturbation makes the internal composition of the system deviate from its normal, “homeostatic” way of functioning. If that deviation becomes too large or happens too suddenly to allow compensation, then autopoiesis breaks down, and the organization dies. As we saw, an autopoietic system modelled as a reaction network defines a dynamical system whose variables are the concentrations of the different molecules. The range in which these variables can change without destroying the autopoietic system defines its *domain of viability*. This viability domain corresponds to the basin of attraction surrounding the attractor, while the attractor itself corresponds to the unperturbed autopoietic process. Indeed, as long as the system remains within this basin, its intrinsic dynamics will make it return to the attractor, and thus automatically restore autopoiesis (Heylighen, 2022).

This ability of a system to recover from perturbations is called *resilience*. It means that if the system is pushed away from its normal regime, it will be able to “bounce back” and recover its essential organization. Resilience has been investigated from a dynamical systems perspective for ecosystems (Meyer, 2016; Walker et al., 2004). Some of the relevant criteria proposed here are the system’s “latitude” (size of the basin of attraction), “resistance” (depth of the basin), and “precariousness” (nearness to the basin’s boundary). Generally speaking, the system will be the more

resilient, the greater the distance between its normal, autopoietic regime (attractor) and the border of the basin, i.e. the greater a perturbation must be to push the system outside of its domain of viability (Heylighen et al., 2022).

The randomly generated example of an autopoietic network, whose reactions are listed in Table 1, was chosen not only because it self-organizes (implying that it has attractors that can be reached from a non-empty basin), but also because it illustrates different levels of resilience. We noted the existence of two attractors, the sets $A_1 = \{a, b, d, e, f, g, i\}$ and $A_2 = \{f, g\}$. This example was generated with an early version of our software that simulates the qualitative dynamics of a reaction network, i.e. the discrete appearance or disappearance of molecules (Heylighen et al., 2015), but not yet the quantitative dynamics, i.e. the continuous variation of molecule concentrations over time (Veloz et al., 2022). The (qualitative) state space of this system is the power set $P(M)$, i.e. the set of all possible combinations of molecules from the 10-element set M . The size of that state space is $2^{10} = 1024$.

The software found that the great majority of these states, namely 927, end up in the attractor A_1 . That means that this attractor has a very large basin (great latitude). It is therefore very resilient: the probability that the combined removal or addition of any number of molecules would make the system end up in a different attractor is only $(1024 - 927) / 1024 = 9.4\%$. The probability that the removal or addition of a single molecule would destroy the organization is actually zero (low precariousness). The reason is that any molecule lost from the organization can be reconstituted by other reactions working on different molecules. The only molecules that can be added (namely c, h, j) are eventually all consumed by reactions, but not produced again. We may conclude that A_1 is a highly resilient organization.

The second organization, A_2 , on the other hand is not resilient at all. It has a basin containing only 2 states out of the 1024 possible ones, namely the states $\{f\}$ and $\{f, g\}$. Practically any molecule change pushes the organization out of its basin and into a different basin, and thus eventually into a different attractor—most likely the attractor A_1 .

In a preceding paper (Heylighen et al., 2022), we have examined some of the general mechanisms that autopoietic systems can use to increase their resilience. These include *negative feedback* (increasing the relative production of molecules that are being depleted), *buffering* (storing reserves of resources that can be recovered when they have been depleted), semi-permeable *membranes* to keep resources in and toxins out, and *degeneracy* (producing the same resources by more than one pathway of reactions, so that if one is interrupted, another one can still function). (It seems to be this last property of degeneracy that explains the exceptional resilience of our example attractor A_1 .) In this paper, we will go into greater depth about those more specific mechanisms for the compensation of perturbations that can be understood as *cognitive*.

Cognition as a resilience mechanism

In a complex environment, such as a natural ecosystem, the variety of potential perturbations is very large. The need for some intelligent action selection mechanism to deal with that complexity can be derived from the famous cybernetic *law of requisite variety*. The law states that the larger the variety of perturbations a system is confronted with, the larger the variety of actions it must be able to perform in order to compensate these perturbations and thus remain within its domain of viability (Ashby, 1958; Heylighen, 1992).

It is worth noting here that a common, stricter formulation of the law, according to which the variety of actions must be—at least—equal to the variety of perturbations, does not hold in general (Heylighen and Joslyn, 2003). One reason is that the goal of the system, in this case the autopoietic regime, still has a variety of potential realizations (what Maturana and Varela call structural changes). Therefore, not all variety must be neutralized. Another reason is that more rudimentary resilience mechanisms, such as degeneracy, buffering and membranes, can absorb a variety of perturbations without requiring any active intervention by the system. Aulin-Ahmavaara (1979) gives the example of a tortoise shell, which absorbs most shocks, thus freeing the tortoise from the need to develop a variety of actions to defend against a variety of predator attacks.

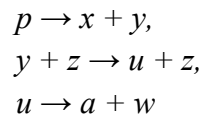
These observations suggest that self-organized autopoietic systems can develop sufficient resilience to survive a variety of conditions without need for something as sophisticated as cognition (Bitbol and Luisi, 2004). That is good news if we wish to understand the origin of life: rudimentary self-maintaining reaction networks seem to be all that is needed to start a process of evolution towards more complex living organisms characterized by cognition (Heylighen et al., 2022).

Still, given that the number of potential perturbations is unlimited, it is useful to increase the variety of actions that can deal with such perturbations. All other things being equal, if two otherwise similar organisms compete, the one with additional variety of actions is more likely to survive, given that sooner or later a perturbation is likely to arise that the first one can counteract, but the second cannot. Therefore, we may assume that as evolution advances, the variety of potential actions will tend to increase (Heylighen, 1999).

However, as variety increases, so does the need for making sure the right action is selected from the repertoire of potential actions. This complementary principle can be formulated as a *law of requisite knowledge* (Heylighen, 1992; Heylighen et al., 2022; Heylighen and Joslyn, 2003): the system must *know* which action to perform in order to compensate a given perturbation. The simplest way to represent such knowledge is what in AI has been called a *condition-action rule* (Holland et al., 1989; Russell and Norvig, 2009). It has the form:

IF a particular condition (perturbation *p*) is sensed,
THEN perform a particular action *a* appropriate for that condition.

As we saw, such a rule is formally equivalent to a reaction, $p \rightarrow a$, with the condition as input and the action (or, more precisely, the new condition resulting from that action) as output. But in a reaction network, it could also be implemented as a more complex process consisting of several, consecutive reactions, involving other molecules, such as:



These intermediate reactions could be seen as “inferences” in a cognitive process that ends in the conclusion a starting from the premise p (Heylighen and Beigi, 2018). Other molecules involved, such as x , y , z , u , and w , may play the role of intermediate steps, enabling conditions or side effects of the cognitive process, which is taking part inside the chemical organization. As long as these do not interfere with the overall autopoietic process that keeps the organization going, they can be seen as merely helping the inference process come to the right conclusion, namely, to produce the action a that will effectively deal with the perturbation p . These molecules may remain present in the autopoietic system in a “dormant” form, i.e. they only become active when a condition arise to which they react, but otherwise are not consumed or produced by the on-going metabolism (Heylighen et al., 2022).

An outside observer may try to understand the dynamics of the autopoietic system by subjecting it to the external “stimulus” p and then noting its (we assume, externally observable) “response” a . However, from this outside perspective, the intermediate reactions remain invisible. Thus, while such an observer may learn to correctly predict that the system will produce a when confronted with p , that observer cannot therefore know which precise processes are taking place inside the system. This brings us to the more general issue of modeling processes to which you do not have direct access.

Is knowledge a representation of reality?

Another, stricter formulation of the law of requisite knowledge is the *good regulator theorem* (Conant and Ashby, 1970), which has also been referred to as the *law of regulatory models*. It states that for a “regulator” to be successful in suppressing all perturbations that make a system deviate from its desired state, its dynamics should include a structure preserving mapping m from the set of potential perturbations P to the set of potential actions A , so that each specific perturbation p is mapped to the specific action a that is appropriate to compensate it:

$$m : P \rightarrow A \text{ so that } p \mapsto m(p) = a$$

The rationale is that systems that respond to perturbations haphazardly, so that sometimes they act in one way, another time in another way, will only succeed part of the time in neutralizing the perturbation. Therefore, they will be worse regulators. Again, this formulation is more restrictive than it needs to be, because sometimes different actions can be equally successful, in the sense that their eventual effect on the perturbation is the same. However, this formulation will help us to clarify an important epistemological issue raised by Maturana and Varela (1980).

Conant and Ashby (1970) interpret this mapping m as a *model* of the system being regulated, in the sense that a one-to-one map establishes a homomorphism between the sets P and A , where P can be interpreted as the externally originating stimuli that make the system deviate from its goal, and A as the repertoire of internal responses that the regulator has access to in order to control those stimuli. In terms of the law of requisite knowledge, this map implements the knowledge that the regulator (which we may here interpret as the cognitive subsystem of the autopoietic system) has on how to select the right action.

This interpretation seems to confirm the traditional epistemology according to which knowledge consists of some internal *representation*, model or map of the external world. According to this epistemology, knowledge is true if the representation is accurate, i.e. if each external phenomenon (stimulus) is mapped onto the corresponding internal concept that represents it (and that determines the eventual response). In this conception, knowledge is acquired simply by perception, which “informs” (organizes or structures) the cognitive system so as to accurately reflect the structure of the environment: external structures are homomorphically mapped to internal ones. Thus, true knowledge is assumed to be an objective reflection of reality (Turchin, 1993).

Maturana and Varela fundamentally reject this objectivist, representationalist philosophy of knowledge (Maturana, 1995; Maturana and Varela, 1980). Their position is similar (though not identical) to what von Glasersfeld (1984) has called *radical constructivism*. That means that knowledge is assumed to be the product of an internal process of construction that helps the organism to survive in its environment. The environment did not “instruct” the organism about how to structure its knowledge; it at most eliminated organisms whose knowledge was inadequate for the purpose of surviving perturbations. Thus, knowledge is intrinsically *subjective*: it depends on the particular way the autopoietic system (subject) is organized.

Maturana (1995) adds that the nervous system, which is supposed to store knowledge in animals and humans, is organizationally closed, just like the autopoietic organism of which it forms part. That means that its activity is essentially self-produced. He notes that *at the level of activation circulating between neurons, there is no distinction between a perception and a hallucination*. Only an outside observer may infer that in the case of a perception there is a correlation between the neural activity and some outside phenomenon, while in the case of a hallucination such correlation is lacking (Maturana, 1995). The organism itself, however, merely knows

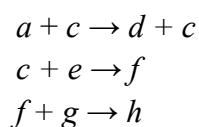
what is happening inside its own network of processes. It therefore cannot develop any objective representation of outside reality.

Formulated in this way, the autopoietic theory of cognition can be easily misinterpreted as *solipsism* (only the inside world of thought is real; there is no outside reality), as an extreme form of *subjectivism* (each subject lives in its own reality, independently of any others), or of *relativism* (any cognitive construction is as good as any other). However, that is not what Maturana and Varela have been saying. We will here not enter into the extensive and still on-going discussion about what their epistemology implies for human knowledge, language and society (see e.g. Bunnell and Riegler, 2022; Proulx, 2008). Instead, we will go back to our reaction network model of autopoiesis in order to explain how knowledge can be a purely internal construction, yet still be an effective “regulator” or “model” in the sense of Conant and Ashby (1970).

The key insight is that the good regulator theorem establishes a mapping from the states of the system to be regulated to the actions that will bring that state back to the desired goal state. If the state of the system is already a goal state, then the corresponding action is simply to do nothing. Therefore, we have assumed that the only states that require specific actions are states that deviate from the goal, i.e. what we called *perturbations*. In the present context, the goal is the autopoietic regime of self-maintenance, i.e. the present attractor of the reaction network (Heylighen, 2022). We also assumed that perturbations originate in the environment, i.e. in molecules and reactions that are not part of the autopoietic organization, but that are still part of our larger reaction network model of on-going processes. That means that we are looking at the overall network from the position of an outside observer. That observer may note how an external event or stimulus (e.g. introduction of a new molecule in the network) leads to an internal action or response (e.g. the initiation of a reaction neutralizing that molecule).

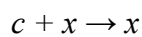
The crucial issue here, which is clarified by the reaction network formalism, is the distinction between inside and outside. “Inside” is defined in chemical organization theory as the set of molecules and reactions that together form a closed, self-maintaining whole. “Outside” then refers to molecules and reactions that do not participate in this self-maintaining organization. However, outside elements may still react with, and thus affect, inside elements. Such reactions are what we called “perturbations” or “stimuli”.

Let us illustrate this with our rudimentary example of an autopoietic network (Fig. 1), namely the circular process: $a \rightarrow b$, $b \rightarrow a$, with inflow $\rightarrow a$, and outflow: $b \rightarrow$. Let us assume that these reactions are part of a larger network that also contains the following reactions:



If now the molecule c is added to the overall network, it will react with the molecule a that is part of the autopoietic cycle. This reaction will consume a and produce d . Therefore, less a will be available to produce b , and thus keep the autopoiesis going. In other words, the addition of c to the system will perturb the autopoietic organization. Happily, there is still an inflow of a , and a conversion from b back to a , so that autopoiesis can continue, albeit at a lower rate. The constant inflow of a and the reduced conversion of a to b together are sufficient to compensate the perturbation to some degree, via a negative feedback mechanism (Heylighen et al., 2022). This is a very simple, non-cognitive form of resilience.

Imagine now that the autopoietic system has acquired a rudimentary form of knowledge, in the form of a reaction:



x plays here the role of a catalyst that neutralizes (consumes) c without being affected itself. We can assume that x is a “dormant” molecule of the autopoietic system, i.e. remaining present in the organization, but in a passive form, not reacting with anything until c appears. Thus, x is kept in the organization’s “memory store”, being retained without changes until it may be needed (Heylighen et al., 2022). Here, x (whose role in a cell may be played by an enzyme that dissolves molecules of type c) embodies or implements a condition-action rule that maps the perturbation c to a (re)action that results in a state where c has been removed, thus re-establishing the normal self-maintaining regime. In this case, we may say that c , the perturbation, belongs to what Maturana and Varela call the *cognitive domain* of the autopoietic system: the system can sense and appropriately respond to the appearance of this perturbation. Its implicit model maps this perturbation to the right action.

Let us now consider the introduction of molecules f and g to the same network of reactions. f does not react with any of the molecules in the autopoietic organization. It may react with g , but, again, the product of that reaction, h , does not interfere with the autopoietic system. Therefore, we can say that none of these molecules is part of the cognitive domain of the initial organization. That autopoietic organization cannot in any way sense whether these molecules are present or absent in its environment. This is similar to the example we discussed of nitrogen, whose presence or absence cannot be perceived by oxygen-breathing animals. Thus, as far as the autopoietic system is concerned, these molecules simply do not exist in its domain of awareness. No representation, map or model of these molecules can be found in the autopoietic organization.

The introduction of respectively c and f to the overall system represent distinct types of events, which are respectively inside the cognitive domain of the organization (mapped to an action) and outside of that domain (ignored). The introduction of e is a less straightforward case, because this molecule does not react with any molecules in the organization, but reacts with c , which is in the organization’s cognitive domain. Indirectly, the reaction of e with c will reduce the

concentration of c , therefore the strength of the perturbation that needs to be compensated, and thus the need for a counteraction. However, the organization does not know whether c is decreasing because it is consumed by a reaction with e , or simply because there is less inflow. At the level of its reactions, there is no difference between both cases. Therefore, we may infer that e remains outside the cognitive domain of the system.

We can provisionally conclude that most events in the autopoietic system's environment will not be mapped to specific actions, and therefore that they remain outside the cognitive domain. Insofar that they enter that domain, the system's "knowledge" or "perception" of these events is limited to the triggering of certain compensatory reactions. These reactions, as we have seen earlier, do not need to have any structural correspondence or isomorphism to the outside reactions that produced that event. Therefore, the reaction network model supports Maturana's and Varela's thesis that knowledge does not presuppose an objective representation of reality: cognition merely serves the subjective need for self-maintenance of the organism, and only "knows" how to deal with whatever interferes directly with that self-maintenance. But the accuracy of that knowledge remains of course crucial: performing the wrong actions (i.e. actions that do not neutralize the perturbation, or that make the system deviate even more from its autopoietic regime) may lead to the system's destruction. What the good regulator theorem (or the more informal law of requisite knowledge) expresses is the need for such accuracy, not the need for an objective representation of external reality.

Cognition as prediction

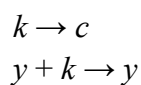
The type of cognition we have discussed up to now is rudimentary: triggering a compensating reaction when a perturbation has made the system deviate from its autopoietic regime. This form of regulation or control relies on *feedback*: suppressing the deviation *after* it has been sensed. However, such after-the-fact reaction runs the risk that it may come too late: the deviation may have grown so strong that it can no longer be repaired. A potentially smarter strategy is to already start suppressing the factors that lead up to the perturbation, *before* the system has actually started to deviate from its autopoietic regime. Such a control strategy may be called *feedforward*: anticipating the effects of external events, and counteracting them before they have affected the system (Heylighen and Joslyn, 2003).

Some of the theorists that investigated autopoiesis and biological autonomy after the pioneering work of Maturana and Varela have suggested that we can only properly speak about cognition when what we called condition-action rules exhibit some form of anticipation (e.g. Bitbol and Luisi, 2004; Bourguine and Stewart, 2004; Di Paolo, 2005; Mossio and Moreno, 2010). This assumption is confirmed by the recently very influential *predictive processing* paradigm in cognitive science, which

sees prediction as the essential function of the brain (e.g. Clark, 2013; Friston, 2018; Hawkins and Blakeslee, 2005; Hohwy, 2013).

However, it is not obvious in what way an autopoietic network can anticipate the effect of events that have not as yet affected its internal reactions. One perspective interprets an action as anticipatory if it suppresses processes moving *towards* the border of the system's domain of viability. That means that the action must be initiated well before the system has reached that border, i.e. before any irreversible damage has occurred (Di Paolo, 2005). However, from a dynamical system perspective, such an action follows automatically from the system's inherent dynamics (Heylighen, 2022). Indeed, the domain of viability here corresponds to the basin of attraction surrounding the attractor (which defines the autopoietic regime). By definition of "basin of attraction", any state in that basin, whether moving towards, sideways, or away from the border, will be led back to the attractor by the system's dynamics—i.e. its network of reactions. Thus, any evolution that would make the autopoietic system more resilient by extending its basin of attraction would thereby also augment its capacity for "anticipatory" action. On the one hand, this suggests that basic aspects of cognition can evolve rather easily, by just extending the range of actions available to neutralize potential perturbations. On the other hand, this does not really clarify what precisely distinguishes "intelligent", "anticipatory" cognition from mere "automatic" application of dynamical rules implemented by reactions.

The reaction network model suggests a different way to characterize anticipatory action (although it as yet does not offer any formal definitions). Let us go back to the example in the previous section of a rudimentary autopoietic cycle consisting of the mutually producing molecules $\{a, b\}$. That cycle is perturbed by the introduction of c , which reacts with a , thus removing it from the cycle. We saw that the autopoietic system could combat this perturbation with the help of some dormant catalyst x that neutralizes c . Assume now that c itself is the product of some other molecule k , which in turn can be neutralized by a similar dormant catalyst y :



The action of y is here anticipating the action of x , by already neutralizing the precursor k of the perturbation c . This looks like a smart move, because if k is removed before its product c is formed, then c cannot damage the essential molecule a . Thus, having y react with k is a form of feedforward regulation: preventing the appearance of a perturbation c that will anyway produce some damage to the organization's metabolism, even if it is almost immediately neutralized by x when it appears (a form of feedback regulation). We may expect that autopoietic systems that have evolved so as to incorporate the anticipatory catalyst y will be fitter than those that only incorporate the corrective catalyst x . Therefore, natural selection can be expected to promote the acquisition of such anticipatory compensation mechanisms.

We can further conceive a more sophisticated anticipatory strategy that is perhaps more typically “cognitive”. There may be no need to neutralize the *in se* innocuous precursor k . However, the appearance of k announces the imminent production of c , which is not innocuous, and needs to be neutralized as quickly as possible by x . This may be achieved by a reaction that is triggered by the presence of k to produce more x , so that sufficient x would be available to neutralize any c that may appear. Thus, k is interpreted by the autopoietic system as a *signal* of a perturbation to come, triggering a number of reactions that prepare the system for neutralizing that perturbation. This strategy also works if k does not participate in the production of the perturbation, but is merely a side effect of some process that eventually leads to the production of c .

Up to now we have been basically considering reactions that produce or consume molecules that directly interact with the autopoietic system’s metabolic network. Yet, the reaction network formalism can just as well describe independent reactions that have an indirect relationship, such as co-occurrence, with the reactions in the metabolism. The products of these reactions, while not interfering with the metabolism, may still be *informative* about the appearance of molecules that do interfere with the metabolism, in the sense that they can be interpreted as warning signals that help the system to prepare a defense against potentially lethal eventualities, but potentially also to get ready to exploit opportunities.

For example, smells are low-concentration, diffuse molecules that do not interact with the basic metabolism, but that may signal the nearness of food or poison, which do interact very directly. The appearance of such challenges requires the appropriate actions, such as approaching the food or evading the poison. Therefore, evolution will tend to make autopoietic systems sensitive to such indirect markers that signal the need for anticipatory action. For example, autopoietic systems may evolve catalysts activated by the smell molecules that set in motion a sequence of reactions for dealing with the anticipated challenges, e.g. producing the enzymes to digest the food or neutralize the toxins.

The evolution of such anticipatory action in regulatory gene networks has recently been demonstrated for bacteria (Freddolino and Tavazoie, 2012; Tagkopoulos et al., 2008). Simulations of such networks indeed evolve the capability to predict perturbations and prepare the metabolism to cope with them. In other words, they learn to react to an imminent perturbation before it has actually happened. Such anticipatory action has also been confirmed empirically. For example, the bacterium *E. coli* normally lives in either a cold, oxygen-rich environment outside the body or in a warm, oxygen-poor environment inside the human gut. When the bacterium is swallowed in the mouth, it will sense an increase in temperature even before it has reached the oxygen-poor environment of the gut. Experiments have shown that if such bacteria are subjected to a warm temperature, they will switch on their anaerobic metabolism in order to cope with the anticipated reduction in oxygen, even when the oxygen supply is still kept high (Tagkopoulos et al., 2008).

A signal may not immediately require anticipatory action, because the challenge it points to will only appear if additional conditions are fulfilled. In this case, the cascade of reactions set in motion by the signaling molecule may merely *prepare* the organism for action (Heylighen and Beigi, 2018), e.g. by already mobilizing the needed resources, while waiting for an additional signal before using these resources to produce the necessary action. Such additional signals can *modulate* the reaction, so that one reaction is triggered in one condition, while a different reaction is triggered in another condition. This allows the autopoietic reaction network to evolve sophisticated “semiotic codes” (Görlich and Dittrich, 2013), where the same signaling molecules can have different pragmatic meanings (i.e. triggering different pathways of action) depending on the context in which they are sensed.

Can we cognize reality?

We have suggested that an autopoietic system is likely to evolve an increasingly sophisticated network of “cognitive” processes that take into account a growing variety of sensed conditions together with their interrelationships. These relationships allow the system to anticipate perturbations and prepare the appropriate actions, thus in practice “predicting” that a sensed signal will be followed by a particular challenge that requires a particular course of action.

The further development of such a cognitive network may eventually take the physical shape of a nervous system, where neurons play the role of catalysts that convert an incoming signal (such as neurotransmitters absorbed) into an outgoing signal (such as an action potential propagated along the neuron’s axon) that is in turn picked up as input by further neurons connected to the first neuron via synapses. We know that in animals, such a neural network has eventually evolved the capacity to *learn*, i.e. adjust the strength of its connections depending on how successful they were in predicting processes or initiating appropriate actions. Thus, we may expect that as the cognitive network evolves and learns, it will become increasingly effective in predicting and controlling outside processes that may affect the internal autopoiesis.

This brings us to an age-old philosophical problem: can our nervous system eventually learn to know reality as it is, i.e. independently of our subjective perspective? Maturana and Varela as well as the radical constructivists (von Glasersfeld, 1984) would see this as a meaningless question. This is one of the most controversial and least understood aspects of the theory of autopoiesis. Therefore, it is worth trying to elucidate it from our reaction network perspective and some more recent work on cognition. According to the theory, an autopoietic system only has access to whatever reacts with its internal metabolism, whether in the form of actual perturbations (such as the ingestion of toxins) or of sensed signals announcing potential perturbations (such as a rotten smell indicating that a piece of food is spoilt and thus likely to contain toxins). It does not have access to whatever outside process

is responsible for these internal reactions. This is the same logic that led Kant to observe that we have no access to the (external) *things-in-themselves*, only to our (internal) perceptions of these. Moreover, the system only truly cares about its internal metabolism, because its overarching goal is survival, i.e. continuing autopoiesis.

Donald Hoffman has proposed a provocative formulation of the evolutionary implications of this logic, which he calls the “Fitness Beats Truth theorem” (Hoffman, 2019; Prakash et al., 2021). His argument (based on a game-theoretic simulation) is that when organisms can choose between an “objective”, “true” representation of reality and one that focuses on the “subjective” aspects most relevant for the organism’s survival, then organisms that choose the latter will win in the evolutionary race. This argument is almost self-evident, given that paying attention to potential fitness benefits will obviously tend to provide more such benefits than paying attention to irrelevant, “objective” properties of the environment. Moreover, as Maturana would note, we anyway do not have access to any objective representation of reality. Hence, what Hoffman is comparing in his evolutionary simulations are merely two models of his own making, none of which can be called objective. Therefore, his “FBT theorem” does not add much to the deeper epistemological issue.

More relevant for the present discussion is the metaphor that Hoffman proposes to understand the subjective nature of the seemingly realistic representations provided by our senses—such as the visual perception of an object. According to his *interface theory of perception* (Hoffman, 2019; Hoffman et al., 2015), an apparently perceived object, such as a tomato or a stone, is rather like an icon on a computer desktop. Such an icon points to a file, while allowing us to manipulate that file. However, it has no intrinsic, structural similarity to that file. There is no isomorphism between the states of the different electronic components that make up the file in the computer’s memory and the pixels on the screen that make up the icon. Yet, we probably would not succeed in achieving anything useful if we had to manipulate the file at the level of its electronic components, while we can easily do that by manipulating the icon. The icon hides the internal complexity of the file, while highlighting its useful functions, thus making it manageable.

In Hoffman’s view, our perceptions of objects are similar: they highlight subjective affordances and potential dangers, while ignoring what the object actually is constituted of. This is consonant with our process ontology, which assumes that apparently simple and stable objects are in reality immensely complex, dynamic and indeterminate networks of interactions between the quantum particles and fields that make up the different atoms, molecules and reactions in that object. Yet, the cognitive reduction of that network of processes to a stable object requires more than a selection of subjectively relevant features. That reduction only works because in the case of rigid objects such as stones (unlike e.g. clouds or waves) the overall organization emerging from these interactions is effectively quite stable. This stability can be explained by the fact that dynamical systems tend to settle down in attractors that

have a much smaller range of variation than the original process. This is similar to how reaction networks settle down in self-maintaining organizations. The relatively invariant properties that emerge from such stabilization (Heylighen, 2001) allow a cognitive system to predict how the object will react to various conditions or actions. Indeed, cognition is only possible in a world that is sufficiently stable so that the rules you have learned in the past can be trusted to still work now and in the future.

The *predictive processing* theory of cognition is inspired by the hierarchical organization of neural networks in the brain. As incoming perceptions are processed by subsequent layers of neurons, their variable, contingent features, such as the changing play of light, perspective, or apparent size relative to the observer, are discarded, while increasingly abstract, invariant patterns, such as shape and absolute size, are extracted (Clark, 2013; Hawkins and Blakeslee, 2005). Thus, the brain is specifically adapted for the recognition of invariant patterns within the constantly varying stimuli that enter through the senses. Our intuitive idea of an “object”, such as a stone, is such an abstraction that groups together the features that remain invariant across different perceptions of the phenomenon.

Cybernetic epistemology (Turchin, 1993) adds that these features should also remain invariant across actions performed by the subject on the phenomenon—such as picking up the stone. Thus, objects can be conceptualized as “eigen-solutions” or “eigen-behaviors” (Stern, 2007; von Foerster, 2003), i.e. those characteristics of sensed phenomena that do not change under their interactions with the subject. Recognizing such invariances allows autopoietic systems to predict and to act more effectively. For example, if I interpret that visual impression of a grey, irregular shape as a real stone (i.e. an invariant object) rather than as a hallucination or a misperception, then I can predict that a similar impression will still be visible when I draw nearer, close and reopen my eyes, or take it in my hand and throw it across the road.

Now, we are ready to discuss whether such internal conceptions correspond to some kind of objective reality. We started by noting that there cannot be an immediate, structure-preserving mapping from outside phenomena to putative internal representations. From the reaction network perspective, reactions both inside and outside the autopoietic system are deeply entangled in different networks that are to some degree autonomous and self-maintaining. Such largely independent networks have only limited points of contact, namely when the product of an outside reaction affects some inside reaction (or vice versa). For example, a photon reflected by a flower may excite a neuron in my retina, or some of the perfume molecules it diffuses may react with chemical sensors in my nose. These points of contact can be conceived as an *interface* between the autopoietic system and its environment: they specify which externally originated conditions the system can sense (e.g. colors, certain smells), and which it can act upon (e.g. grasp the flower with your fingers). These limited interactions, however, function as signals that set in motion complex cognitive processes of anticipation and action. These processes may recognize invariant patterns within the sequence of sensed signals, and use these to summarize the most relevant

and dependable features of the environment. They do that by creating the internal equivalent of an icon in an interface, i.e. a simple and intuitive pointer (e.g. the conception of a “flower”) indicating something that is intrinsically complex and dynamic, but that is sufficiently stable so that you can dependably predict or manipulate some of its aspects.

Again, does such an internally conceived object, such as our conception of a flower or a stone, depict “reality”? Not in the sense that it would be isomorphic to the network of processes that produced the sensations from which it is derived. The network of organic, metabolic processes that make up the flower or the rigid crystalline bonds between the atoms that make up the stone have no equivalent in my neural conception of these objects.

However, my conception that there is a stone on the table in front of me is realistic in the more pragmatic sense that different people looking from different perspectives at that table will all agree that there is a stone there, and that I will still perceive that stone if I close my eyes and then open them again. In practice, we distinguish between perception and hallucination by the degree of invariance of the corresponding sensations (Bonsack, 1977; Heylighen, 1997). If I merely dreamt or imagined that there was a stone on the table, then I would not be able to see that stone at a later moment, and neither would other people. The larger the number of perspectives, modes of perception, or manipulations that a phenomenon can undergo without losing its characteristic features, the more “real” we will consider it to be.

In conclusion, what we call “reality” is largely constructed by autopoietic, cognitive systems, individually or collectively, as a way to organize and make sense of the interactions in which they engage. However, for such a cognitive construction to have any benefit for autopoiesis it should as much as possible capture the kind of invariances or regularities in the flow of processes that allow dependable predictions and actions. If science is considered to be our best tool for understanding the world in which we live, that is because science is systematically searching for invariances in observations, formulating hypotheses about the underlying patterns, and then performing experiments to test the predictions implied by these hypotheses. If the predictions turn out to be wrong, then the hypothesis is abandoned, and a better one is searched for.

However, that does not mean that a model that makes good predictions would depict reality as it is, independently of any observer. First, there will still be an infinite number of processes that the model cannot predict, so any model is highly limited. Second, different models or theories of the same phenomena (e.g. different interpretations of quantum mechanics) can often make the same predictions, while disagreeing about what these phenomena actually are constituted of. For example, is an electron a particle, a wave, an excitation of the electromagnetic field, a superstring, a process, a probability distribution, a type of consciousness, or a Platonic form? From an operational point of view, the differences between these interpretations only matter insofar that they lead to different testable predictions. If they do not, the choice of which one the observer assumes to be true will depend more on taste, intuition or

aesthetics than on any underlying reality. That does not mean that they do not matter: after all, intuition is a powerful guide in making sense of complex phenomena and thus inferring non-obvious implications of the given data. However, we must be aware that our intuition is shaped by our particular biological, personal and cultural history, and that there are likely to be other intelligent agents (i.e. autopoietic systems), human or non-human, that have different intuitions and therefore different conceptions of reality...

Conclusion

This paper has revisited some of the classic ideas of Maturana and Varela on autopoiesis and cognition, albeit from a new perspective. This perspective starts from a process ontology, in which the world is conceived as a network of processes rather than as a collection of static objects (Heylighen, 2023). This ontology is implicit in the original definition of an autopoietic system (Meincke, 2018) as a network of processes that regenerates itself. The power of the concept of autopoiesis is precisely that it explains how an organism can appear as a stable, well-defined object, yet consist of a metabolic cycle in which components are continuously being broken down and reconstituted again. Such an entanglement of stability and change is counterintuitive, complex, and confusing. That may explain in part why the work of Maturana and Varela has often been misunderstood, and why the notion of autopoiesis has made less of an impact in biology than it deserves.

Our approach has been to clarify the concept by proposing a simple formalism for modeling such entangled processes: *reaction networks*. A reaction represents an elementary process in which certain conditions are produced, consumed, or transformed into other conditions. In the context of biological metabolism, the most obvious interpretation of the conditions is as the presence of particular molecules. Yet, in fact, the formalism can represent any kind of component, structure or context that can be transformed into another one. What counts is how the production of a condition by one reaction may trigger one or more other reactions that produce further conditions, thus setting in motion a cascade of productions and consumptions propagating within the overall reaction network (Heylighen, 2023).

In some cases, this cascade settles into a self-regenerating cycle, i.e. a subnetwork of reactions that trigger each other, so that the activity does not die down, but is sustained indefinitely. This abstract model of an autopoietic system has been called a *chemical organization* (Dittrich and Fenizio, 2007). The requirement for a set of molecules and reactions to be a chemical organization is that it is *closed*, in the sense that the reactions do not produce molecules outside the initial set, and *self-maintaining* in the sense that all molecules in the initial set consumed by some reactions are produced again by other reactions. That means that the set of molecules is invariant under the reactions that act on them. Such self-regenerating reaction networks are attractors of the reaction dynamics. Therefore, they tend to self-

organize. By simulating the dynamics of randomly generated networks, our research group is presently exploring what kinds of conditions are most conducive to the emergence of non-trivial, autopoietic networks (Veloz et al., 2022).

To model the origin of autopoiesis in a realistic environment (Heylighen et al., 2022), however, such simulations must also include a variety of external processes that may perturb the system. To survive in such precarious circumstances, the system must be *resilient*, i.e. able to bounce back after it has been disturbed. In a preceding paper (Heylighen et al., 2022), we reviewed a number of properties that can make an autopoietic network more resilient. These include degeneracy, a protective membrane, buffering, and regulation.

The regulation strategy, as demanded by the cybernetic laws of requisite variety and of regulatory models, entails *cognition*: the system must “know” which action to perform in order to compensate which perturbation. Such knowledge can be expressed mathematically as a homomorphism mapping the set of perturbations to the set of compensatory actions. Such a mapping can be interpreted as an implicit *model* of the perturbations that the system can experience (Conant and Ashby, 1970). However, it cannot be interpreted as a representation of the external reality where these perturbations originated. The reason is that the autopoietic system only has access to its own processes and the way these are perturbed, not to any external causes of these perturbations. Therefore, the epistemology implied by the theory of autopoiesis is constructivist, and not representationalist, realist or objectivist.

The fact that the knowledge held by an autopoietic system is fundamentally subjective does not mean that any model is as good as any other. As organisms evolve, they must become more resilient or adaptive, in the sense that they acquire a greater variety of reactions that allow them to cope with a greater variety of sensed, and eventually anticipated, perturbations. Such anticipation can be achieved through compensatory reactions that are triggered by molecules that merely signal an imminent perturbation. That allows the organism to already prepare an appropriate course of action *before* it is hit by the perturbation.

Thus, next to the selection of appropriate responses, a fundamental function of cognition is the accurate *prediction* of events. That seems to imply that the internal cognitive model should objectively reflect some external reality. However, the only thing we can conclude from the reaction network formalism is that some internal mechanism adequately anticipates internal sensations resulting from putative external processes. In other words, for an autopoietic system, reality remains a *black box*: it may be able to determine which inputs are followed by which outputs, but it cannot therefore discern the intermediate processes taking place inside the box.

Still, the fact that reliable prediction is possible suggests that these processes must be characterized by some degree of invariance, so that patterns learned in the past remain applicable at later moments. That invariance also makes it possible for different observations using different perspectives to come to the same conclusion about what has been observed. Thus, in a pragmatic sense, observers can agree about

whether an apparent perception refers to some “real” (i.e. invariant) phenomenon rather than being a hallucination.

That does not imply that good models must be isomorphic to the processes whose outcomes they predict. This is confirmed by chemical organization theory, which demonstrates how apparently simple, self-maintaining sets of molecules can emerge from complicated, branching cycles of reactions (such as the one in Table 1). Thus, we come back to the original insight of Maturana and Varela, namely that apparently simple and stable units, such as biological individuals, should properly be understood as complex, self-producing networks of processes. The moral of the story is then, perhaps, that the world is much more complex, interconnected, and dynamic than the traditional, object-based ontology assumes, but that reasoning in terms of objects can still be a useful strategy—as long as these objects are understood as mere icons pointing to relatively invariant organizations within an immense network of ever-varying processes.

The main contribution of the reaction network model is that it proposes an explicit, formal mechanism for how such invariant organizations can emerge from the surrounding processes. In future research we hope to elucidate some more of the details of that mechanism and its implications for the origin and evolution of life and cognition through further conceptual analysis and computer simulation of reaction networks (Heylighen et al., 2022; Veloz et al., 2022).

Acknowledgments

This research was funded by the John Templeton Foundation as part of the project “The Origins of Goal-Directedness” (grant ID61733). We thank our VUB colleagues collaborating on this project (Veloz et al., 2022) for many inspiring discussions on the concepts presented here.

Declarations of interest:

none

References

- Ashby, W.R., 1958. Requisite variety and its implications for the control of complex systems. *Cybernetica* 1, 83–99.
- Aulin-Ahmavaara, A.Y., 1979. The law of requisite hierarchy. *Kybernetes* 8, 259–266.
- Beer, R.D., 2020. An Investigation into the Origin of Autopoiesis. *Artif. Life* 26, 5–22. https://doi.org/10.1162/artl_a_00307
- Bergson, H., 1911. *Creative evolution*. Henry Holt.
- Bich, L., Etxeberria, A., 2013. Systems, Autopoietic, in: Dubitzky, W., Wolkenhauer, O., Cho, K.-H., Yokota, H. (Eds.), *Encyclopedia of Systems Biology*. Springer, New York, NY, pp. 2110–2114. https://doi.org/10.1007/978-1-4419-9863-7_83
- Bickhard, M.H., 2011a. Systems and Process Metaphysics, in: Hooker, C. (Ed.), *Philosophy of Complex Systems, Handbook of the Philosophy of Science*. North-Holland, Amsterdam, pp. 91–104. <https://doi.org/10.1016/B978-0-444-52076-0.50002-X>
- Bickhard, M.H., 2011b. Some Consequences (and Enablings) of Process Metaphysics. *Axiomathes* 21, 3–32. <https://doi.org/10.1007/s10516-010-9130-z>
- Bitbol, M., Luisi, P.L., 2004. Autopoiesis with or without cognition: defining life at its edge. *J. R. Soc. Interface* 1, 99–107.
- Bonsack, F., 1977. Invariance as a Criterion of Reality. *Dialectica* 31, 313–331.
- Bourgine, P., Stewart, J., 2004. Autopoiesis and cognition. *Artif. Life* 10, 327–345.
- Clark, A., 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204.
- Conant, R.C., Ashby, W.R., 1970. Every good regulator of a system must be a model of that system. *Int. J. Syst. Sci.* 1, 89–97.
- Deacon, T.W., Srivastava, A., Bacigalupi, J.A., 2014. The transition from constraint to regulation at the origin of life. *Front Biosci* 19, 945–57.
- Di Paolo, E.A., 2005. Autopoiesis, Adaptivity, Teleology, Agency. *Phenomenol. Cogn. Sci.* 4, 429–452. <https://doi.org/10.1007/s11097-005-9002-y>
- Dittrich, P., Fenizio, P.S. di, 2007. Chemical Organisation Theory. *Bull. Math. Biol.* 69, 1199–1231. <https://doi.org/10.1007/s11538-006-9130-8>
- Dittrich, P., Winter, L., 2008. Chemical organizations in a toy model of the political system. *Adv. Complex Syst.* 11, 609. <https://doi.org/10.1142/S0219525908001878>
- Finkelstein, D., 1973. A Process Conception of Nature, in: Mehra, J. (Ed.), *The Physicist's Conception of Nature*. Springer Netherlands, Dordrecht, pp. 709–713. https://doi.org/10.1007/978-94-010-2602-4_38
- Flamm, C., Stadler, B.M.R., Stadler, P.F., 2015. Generalized Topologies: Hypergraphs, Chemical Reactions, and Biological Evolution, in: Basak, S.C., Restrepo, G., Villaveces, J.L. (Eds.), *Advances in Mathematical Chemistry and Applications*. Bentham Science Publishers, pp. 300–328. <https://doi.org/10.1016/B978-1-68108-053-6.50013-2>
- Freddolino, P.L., Tavazoie, S., 2012. Beyond Homeostasis: A Predictive-Dynamic Framework for Understanding Cellular Behavior. *Annu. Rev. Cell Dev. Biol.* 28, 363–384. <https://doi.org/10.1146/annurev-cellbio-092910-154129>
- Friston, K., 2018. Does predictive coding have a future? *Nat. Neurosci.* 21, 1019–1021.

- Görlich, D., Dittrich, P., 2013. Molecular Codes in Biological and Chemical Reaction Networks. PLOS ONE 8, e54694. <https://doi.org/10.1371/journal.pone.0054694>
- Hättich, F., 2004. Quantum Processes: A Whiteheadian Interpretation of Quantum Field Theory. Agenda.
- Hawkins, J., Blakeslee, S., 2005. On Intelligence: How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines, Reprint edition. ed. St. Martin's Griffin, New York, NY.
- Heylighen, F., 2023. Relational agency: a new ontology for co-evolving systems, in: Corning, P., Kauffman, S.A., Noble, D., Shapi, J.A., Vane-Wright, R.I., Pross, A. (Eds.), Evolution 'On Purpose': Teleonomy in Living Systems, Vienna Series in Theoretical Biology. MIT Press, pp. 79–104.
- Heylighen, F., 2022. The meaning and origin of goal-directedness: a dynamical systems perspective. Biol. J. Linn. Soc. blac060. <https://doi.org/10.1093/biolinnean/blac060>
- Heylighen, F., 2011. Self-organization of complex, intelligent systems: an action ontology for transdisciplinary integration. Integral Rev.
- Heylighen, F., 1999. The growth of structural and functional complexity during evolution, in: Heylighen, F., Bollen, J., Riegler, A. (Eds.), The Evolution of Complexity. Springer, pp. 17–44.
- Heylighen, F., 1997. Objective, subjective and intersubjective selectors of knowledge. Evol. Cogn. 3, 63–67.
- Heylighen, F., 1992. Principles of Systems and Cybernetics: an evolutionary perspective, in: Cybernetics and Systems' 92. pp. 3–10.
- Heylighen, F., 1990. Relational Closure: a mathematical concept for distinction-making and complexity analysis, in: In: Cybernetics and Systems '90, R. Trappl (Ed.), (World Science, Singapore), p. 335-342.
- Heylighen, F., Beigi, S., 2018. Mind outside Brain: a radically non-dualist foundation for distributed cognition, in: Carter, J.A., Clark, A., Kallestrup, J., Palermos, S.O., Pritchard, D. (Eds.), Socially Extended Epistemology. Oxford University Press, Oxford. <https://doi.org/10.1093/oso/9780198801764.003.0005>
- Heylighen, F., Beigi, S., Busseniers, E., 2022. The role of self-maintaining resilient reaction networks in the origin and evolution of life. Biosystems 219, 104720. <https://doi.org/10.1016/j.biosystems.2022.104720>
- Heylighen, F., Beigi, S., Veloz, T., 2015. Chemical Organization Theory as a modeling framework for self-organization, autopoiesis and resilience (ECCO Working Papers No. 2015– 01).
- Heylighen, F., Joslyn, C., 2003. Cybernetics and Second-Order Cybernetics, in: Robert A. Meyers (Ed.), Encyclopedia of Physical Science and Technology. Academic Press, New York, pp. 155–169.
- Hoffman, D., 2019. The Case Against Reality: Why Evolution Hid the Truth from Our Eyes. W. W. Norton & Company, New York.
- Hoffman, D.D., Singh, M., Prakash, C., 2015. The Interface Theory of Perception. Psychon. Bull. Rev. 22, 1480–1506. <https://doi.org/10.3758/s13423-015-0890-8>
- Hohwy, J., 2013. The Predictive Mind. Oxford University Press.
- Holland, J.H., Holyoak, K.J., Nisbett, R.E., Thagard, P.R., 1989. Induction: processes of inference, learning, and discovery. MIT Press.
- Hordijk, W., Steel, M., Dittrich, P., 2018. Autocatalytic sets and chemical organizations:

- Modeling self-sustaining reaction networks at the origin of life. *New J. Phys.* 20, 015011.
- Kuhlmann, M., 2000. Processes as objects of quantum field theory. *Faye A1* 365–388.
- Luhmann, N., 1986. The autopoiesis of social systems, in: Geyer F. & van der Zouwen J. (Ed.), *Sociocybernetic Paradoxes*. Sage, London, pp. 172–192.
- Luisi, P.L., 2003. Autopoiesis: a review and a reappraisal. *Naturwissenschaften* 90, 49–59.
- Luisi, P.L., 1994. The chemical implementation of autopoiesis, in: *Self-Production of Supramolecular Structures*. Springer, pp. 179–197.
- Maturana, H.R., 1995. Cognitive Strategies, in: Von Foerster, H. (Ed.), *Cybernetics of Cybernetics*. Future Systems, pp. 457–469.
- Maturana, H.R., Varela, F.J., 1980. *Autopoiesis and Cognition: The Realization of the Living*. D Reidel Pub Co.
- McMullin, B., Varela, F.J., 1997. Rediscovering Computational Autopoiesis, in: Husbands, P., Harvey, I. (Eds.), *Fourth European Conference on Artificial Life*. Presented at the Fourth European Conference on Artificial Life, MIT Press.
- Meincke, A.S., 2018. Autopoiesis, biological autonomy and the process view of life. *Eur. J. Philos. Sci.* 9, 5. <https://doi.org/10.1007/s13194-018-0228-2>
- Meyer, K., 2016. A Mathematical Review of Resilience in Ecology. *Nat. Resour. Model.* 29, 339–352. <https://doi.org/10.1111/nrm.12097>
- Montévil, M., Mossio, M., 2015. Biological organisation as closure of constraints. *J. Theor. Biol.* 372, 179–191. <https://doi.org/10.1016/j.jtbi.2015.02.029>
- Mossio, M., Moreno, A., 2010. Organisational Closure in Biological Organisms. *Hist. Philos. Life Sci.* 32, 269–288.
- Peter, S., Dittrich, P., 2011. On the Relation between Organizations and Limit Sets in Chemical Reaction Systems. *Adv. Complex Syst.* 14, 77–96. <https://doi.org/10.1142/S0219525911002895>
- Peter, S., Ibrahim, B., Dittrich, P., 2021. Linking Network Structure and Dynamics to Describe the Set of Persistent Species in Reaction Diffusion Systems. *SIAM J. Appl. Dyn. Syst.* 20, 2037–2076. <https://doi.org/10.1137/21M1396708>
- Prakash, C., Stephens, K.D., Hoffman, D.D., Singh, M., Fields, C., 2021. Fitness Beats Truth in the Evolution of Perception. *Acta Biotheor.* 69, 319–341. <https://doi.org/10.1007/s10441-020-09400-0>
- Pressman, A., Blanco, C., Chen, I.A., 2015. The RNA World as a Model System to Study the Origin of Life. *Curr. Biol.* 25, R953–R963. <https://doi.org/10.1016/j.cub.2015.06.016>
- Reichel, A., 2011. Snakes all the Way Down: Varela's Calculus for Self-Reference and the Praxis of Paradise. *Syst. Res. Behav. Sci.* 28, 646–662. <https://doi.org/10.1002/sres.1105>
- Rescher, N., 2000. *Process philosophy: a survey of basic issues*. Univ of Pittsburgh Pr.
- Russell, S.J., Norvig, P., 2009. *Artificial intelligence: a modern approach*. Prentice hall.
- Seibt, J., 2022. Process Philosophy, in: Zalta, E.N. (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
- Stern, J.M., 2007. Cognitive constructivism, eigen-solutions, and sharp statistical hypotheses. *Cybern. Hum. Knowing* 14.
- Tagkopoulos, I., Liu, Y.-C., Tavazoie, S., 2008. Predictive Behavior Within Microbial Genetic Networks. *Science* 320, 1313–1317. <https://doi.org/10.1126/science.1154456>
- Teilhard de Chardin, P., 1959. *The phenomenon of man*. Collins London.

Turchin, V., 1993. On Cybernetic Epistemology. *Syst. Res.* 10, 3–3.

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

Varela, F.J., Maturana, H.R., Uribe, R., 1974. Autopoiesis: the organization of living systems, its characterization and a model. *Biosystems* 5, 187–196.

Veloz, T., Maldonado, P., Busseniens, E., Bassi, A., Beigi, S., Lenartowicz, M., Heylighen, F., 2022. Towards an Analytic Framework for System Resilience Based on Reaction Networks. *Complexity* 2022, e9944562. <https://doi.org/10.1155/2022/9944562>

Veloz, T., Razeto-Barry, P., 2017. Reaction networks as a language for systemic modeling: Fundamentals and examples. *Systems* 5, 11. <https://doi.org/10.3390/systems5010011>

von Foerster, H. (Ed.), 2003. Objects: Tokens for (Eigen-)Behaviors, in: *Understanding Understanding: Essays on Cybernetics and Cognition*. Springer, New York, NY, pp. 261–271. https://doi.org/10.1007/0-387-21722-3_11

Von Glasersfeld, E., 1984. An introduction to radical constructivism. *Inven. Real.* 17–40.

Walker, B., Holling, C.S., Carpenter, S.R., Kinzig, A., 2004. Resilience, adaptability and transformability in social–ecological systems. *Ecol. Soc.* 9, 5.

Walsh, D.M., 2018. Objectcy and agency: Towards a methodological vitalism, in: *Everything Flows*. Oxford University Press.

Whitehead, A.N., 1978. *Process and Reality: An Essay in Cosmology* Ed. David Ray Griffin and Donald W. Sherburne. Free Press, New York.