Chemical Organization Theory as a universal modeling framework for interaction, self-organization, and autopoiesis

Francis Heylighen, Shima Beigi and Tomas Veloz
Global Brain Institute
Vrije Universiteit Brussel

Abstract: xxxx

Introduction

Complex adaptive systems are systems consisting of many interacting agents that exhibit some degree of self-organization: coherent patterns of forms of organization spontaneously emerge out of the network of interactions. Most of the phenomena we are confronted with in real life are such complex adaptive systems: people, organisms, cells, societies, ecosystems, markets, communities, languages, cultures… Great progress has been made in understanding the dynamics of such systems by means of multi-agent computer simulations. However, on the more general theoretical level, our understanding of self-organization and adaptation remains rather superficial and fragmented.

Part of the reason is that the conceptual and mathematical building blocks of our theories are poorly fitted to describe emergence and interaction. Scientific models start by analyzing a system into its static components together with the properties in which these components can vary. The values of these different variables then define the state of the system. The different possible states determine a predefined state space. Only after introducing all these static elements can we start to describe the evolution of that system as a predetermined trajectory in state space. This approach makes it very difficult to understand the kind of fundamental changes that lead to the emergence of new components, properties or systems.

An alternative approach is to start from a process metaphysics (Rescher, 1996; Whitehead, 1978) or action ontology (Heylighen, 2011; V. F. Turchin, 1993). Such a philosophy assumes that reality is not constituted out of static “things”, but out of dynamic processes or actions. The problem until now was to represent such processes in a way that is precise, simple and concrete. As noted, traditional scientific formalisms model processes as transitions between states. In addition to an a priori specification of the state space, that requires a system of equations, laws or dynamics that specify why and how a particular state is mapped onto a particular other state.

This paper wishes to introduce an alternative formalization of processes, namely the reaction networks used in so-called Chemical Organization Theory (COT) (Dittrich & Fenizio, 2007; Dittrich & Winter, 2008; Peter, Veloz, & Dittrich, 2011). In COT, the relation between states and processes is
turned upside down. The dynamics is primary, in the form of the “reactions”, which are the most fundamental elements of a reaction system. States only appear in a second stage, as the changing concentrations of the “molecules” that the reactions are processing into other molecules. The molecules therefore are not static objects, but merely the raw materials that are constantly being produced, consumed, and recreated by the reactions. In that sense, COT seems to be the first formalization of a process ontology that is both practical and fundamental.

The fundamental character is shown by the fact that COT can describe systems and processes in any discipline—from elementary particle reactions via systems biology and cognitive science to the political organization of society. Its particular strength is that it provides an elegant mathematical method to define and construct organizations, i.e. self-sustaining networks of interactions within a larger network of potential interactions. As such, it is eminently suited to describe self-organization, autopoiesis, individuation, sustainability, resilience, and the emergence of complex, adaptive systems out of simpler components.

Next to its deep philosophical foundation, COT derives its power from its concreteness and simplicity: basically, you can represent any process in the real world as a combination of reactions between suitably chosen “molecules”, and then start analyzing the resulting reaction system for self-maintenance, closure, and other observable properties. Moreover, COT models are eminently modular: it is trivial to add or to remove molecules or reactions from an existing model, and (somewhat less trivial) see what effect that has on the emerging organizations and other properties. This makes it possible to model systems of great complexity, where you start with a simple model in order to get an intuitive grasp of what is going on, and then gradually add more detail and sophistication in order to achieve a more realistic representation.

Finally, COT focuses on what are the truly most important questions about a complex adaptive system: in how far is it sustainable and resilient, i.e. able to maintain itself both autonomously and in the face of external perturbations? In how far does it grow, remain the same, or perhaps diminish and decay? If it is perturbed to such a degree that it cannot maintain its present organization, which new type of organization is it likely to evolve into? How do its components and processes co-evolve, mutually adapt, and become coordinated into a synergetic whole? In summary, how does it self-organize into a robust, coherent whole?

The latter is perhaps the most important question in the whole of science and philosophy, and their applications to society. Practically all the phenomena we are confronted with—including matter, life, ecosystems, societies, markets, networks and minds—are the result of self-organization producing complex wholes out of simpler components. Any general theory that would help us to understand, model and control that process is likely to revolutionize our worldview, while opening up an endless variety of concrete applications.

The present paper wishes to make the case that COT, together with its future extensions, provides an extremely promising foundation for such a general theory. It will do that first by pointing out how the COT formalisms avoids the pitfalls of earlier approaches, then by offering a preliminary survey of existing and potential applications of COT in a wide variety of domains. It will do this in a simple, non-technical way, emphasizing the basic formalism and the core new insights, while avoiding some of the (relatively) more complex mathematical techniques required for a full implementation of COT.
Reaction networks

As its name implies, the COT formalism is inspired by chemistry, and the way it describes how chemical reactions combine molecules into new molecules. Therefore, it has inherited much of its terminology from chemistry. In order to widen its appeal and to convince other scientists of its potential for transdisciplinary unification, it may be wiser to replace some specifically chemical terms by more broadly applicable ones. The present paper will therefore replace some COT terms by new terms—however, while clearly pointing out the changes.

The basis of a COT model is a reaction network. It consists of two types of entities, which we will call resources (“molecules”, “molecular species”, or “species” in the conventional COT formulation) and reactions. A resource is an abstract representation of a specific kind of substance, entity, or, most generally, distinguishable phenomenon. Some examples of resources are particular molecules, elementary particles, biological species, products, objects, tools, messages, words, ideas, or statements. All the resources in the model are assumed to be available in some shared “workspace” or container, which in COT is called the “reaction vessel”, but which we will simply call the medium. This joint presence allows them to interact or react with each other. Reactions denote elementary processes that create or destroy resources. They typically produce combinations of new resources out of combinations of existing resources. However, the simplest reactions just create or destroy a single resource.

Formally, a reaction network is defined by the 2-tuple \(<M, R>\), where \(M = \{a, b, c, \ldots\}\) is the set of resources, and \(R \subseteq P(M) \times P(M)\) is the set of reactions, where \(P(M)\) denotes the power set (i.e. the set of all subsets) of \(M\). Each reaction transforms a particular subset \(X\) of \(M\) into another subset \(Y\) of \(M\):

\[
r: X \rightarrow Y: \{x_1, x_2, \ldots | x_i \in M\} \rightarrow \{y_1, y_2, \ldots | y_i \in M\}
\]

We will call \(X\) the input set and \(Y\) the output set of \(r\), and denote them respectively \(I(r)\) and \(O(r)\). We will call the elements of \(I(r)\) the reactants of \(r\), and the elements of \(O(r)\) its products. Borrowing the chemical notation for reactions, a reaction is conventionally written as:

\[
r: x_1 + x_2 + \ldots \rightarrow y_1 + y_2 + \ldots
\]

The “+” symbol here represents a conjunction of the necessary resources: \(x_1, x_2, \ldots\) all need to be simultaneously present in \(I(r)\) for the reaction to take place, while the reaction simultaneously produces \(y_1, y_2, \ldots\).

Note that in traditional COT it is assumed that \(I(r)\) and \(O(r)\) are multisets. This means that the same element \(x_i\) can occur more than once (say \(n_i\) times) in \(I(r)\) This is necessary to describe reactions of the form:

\[
2a + b \rightarrow 3c + d,
\]
or more generally:

\[ n_1 x_1 + n_2 x_2 + \ldots \rightarrow m_1 y_1 + m_2 y_2 + \ldots \text{ with } n_i, m_j \in \mathbb{N} \]

For simplicity, we will here work only with ordinary sets, i.e. resources that only occur once in a reaction. If necessary, reactions with more than one copy of a resource can be represented as a sequence of substages, each adding just one instance of a resource. E.g. \(2a + b \rightarrow e\) could be rewritten as \(a + b \rightarrow x, x + a \rightarrow e\).

![Fig. a depiction of the reaction \(r: a + b + c \rightarrow d + e\)](image)

**Reaction Networks vs. Traditional Networks**

The combined system \(<M, R>\) forms a network because the resources in \(M\) are linked to each other by the reactions in \(R\) that transform the ones into the others. But this is not a traditional network (i.e. a directed graph), in which a link connects a single element (“node”, “vertex”) \(x\) to a single element \(y\). A reaction connects a set \(X\) of elements to a set \(Y\) of elements. In mathematics, a network with this property is called a directed hypergraph (Gallo, Longo, Pallottino, & Nguyen, 1993). This appears to be the essential generalization that gives reaction networks their power with respect to traditional network models. Let us try to explain how that happens.

A traditional network consists of nodes \(N\) and links \(L\), with \(L \subseteq N \times N\). Thus, \(<L, N>\) is a reaction network, but where the reactions \(r \in L\) are limited to one input and one output:

\[ r: x \rightarrow y, \text{ with } x, y \in N \]

A general reaction network provides much more richness and flexibility because it allows combinations of inputs to produce combinations of outputs, opening up an exponentially wider range of interacting processes. In a traditional network, the only way processes can “interact” is by sharing input or output nodes, e.g.
Here an initial state $x$ can lead to $y$ and/or to $z$ via respectively $r_1$ and $r_2$. This immediately creates an ambiguity: are $r_1$ and $r_2$ both taking place, producing $y$ and $z$ simultaneously? Or does the process make a choice between $r_1$ and $r_2$, ending up in either $y$ or $z$? Similarly, $y$ can be produced via $r_1$ and/or $r_3$ from $x$ and/or $u$. Do we need both $x$ and $u$, or is one of them sufficient to produce $y$? The problem is that in traditional networks there is no way to distinguish between conjunction (“AND”) and disjunction (“OR”) of nodes and links. Next to juxtaposition of links/reactions, there simply is no operator to express a different type of combination.

In reaction networks, we have an additional operator, denoted by “+”, that plays the role of the conjunction. The juxtaposition of reactions plays the role of the disjunction. Consider the following reactions:

\[
\begin{align*}
  r_4 &: x + u \rightarrow y \\
  r_5 &: w \rightarrow y 
\end{align*}
\]

This expresses unambiguously that in order to produce $y$ we need either ($x$ AND $u$), OR $w$. Now, different processes can interact in many ways to produce complex organizations, as we will see in further sections.

**Reaction networks and propositional logic**

Interpretation of COT operators in terms of conjunction and disjunction points us towards formal, Boolean logic, where propositions can be combined in terms of these operators, together with the operators of implication and negation. Implication is naturally expressed by the “$\rightarrow$” operator. This directly suggests the logical formalism of Horn clauses (Chandra & Harel, 1985). These have the following form:

\[ a \land b \land \ldots \rightarrow x \]

This is to be read as “if $a$ and $b$ and … are true, then $x$ is true”, or “$x$ can be derived from the conjunction of $a$, $b$, …”. The translation in terms of reaction networks requires a qualification, though, which is that if you derive the new proposition $x$ from the conjunction of propositions on the input side of the inference, then the propositions on that side remain actual (they are not destroyed by the process). They therefore should properly be listed on the output side as well. This gives us the straightforward COT translation of a Horn clause:

\[ a + b + \ldots \rightarrow x + a + b + \ldots \]
Note that in this interpretation, logical inference is a special type of reaction, namely one in which no “resources” ever get consumed: inferences can only add true propositions to our knowledge, they cannot remove any. This is why logic is inherently static: nothing really changes by making logical inferences; at most we become aware of additional statements that were already true implicitly, but had not been proven yet. That is the fundamental reason why all attempts to derive process, actions, dynamics or time from logic are bound to fail, in spite of a plethora of formalization attempts such as “dialectical logics”, “process logics” or “dynamic logics” (see e.g. Harel, Kozen, & Parikh, 1982; Ilyenkov, 2008; Van Benthem, 2007). Insofar that these logics describe genuine changes in the state of the world, they have left the domain of logic proper and entered the domain of dynamics, which is more properly described by a formalism such as reaction networks.

The last fundamental operator of logic, negation, does not seem to have any equivalent in the COT framework. However, we can introduce it by applying the same convention that you use when moving variables from one side of an equation to the other side: the variable changes sign, from positive (“+”) to negative (“−”). For example:

\[ r6: a + b \rightarrow c + d \]

could be rewritten as:

\[ a \rightarrow c + d - b \]

The interpretation is that r6 consumes b, and therefore it produces the absence (“negation”) of b, which we write as “−b”. In logic we have the similar principle of contraposition, which states that propositions turn into their negations when moved to the other side of an implication:

\[ a \rightarrow b \iff \neg b \rightarrow \neg a \]

To express the negation of a single resource, we may note that reactions can have the empty set (\(\emptyset\)) as their input or output. The production of a out of nothing (“affirmation”) can then be written as:

\[ \emptyset \rightarrow a, \text{ or more simply: } \rightarrow a \]

The pure consumption or elimination (“negation”) of a is then written as:

\[ \rightarrow \]

According to the rule of changing signs when moving to the other side of the arrow, this becomes:

\[ \rightarrow \neg a \]

This is in turn equivalent to the affirmation of −a (“not a”).
Yet another way to grasp the meaning of the negation operator is to start from the “trivial” reaction, which merely asserts that if $a$ then $a$:

$$a \rightarrow a$$

Moving $a$ to the other side then produces the following two reactions:

$$a + (-a) = a - a \rightarrow$$

$$\rightarrow a - a$$

An intuitive interpretation in terms of processes is that $a$ and $-a$ are like a particle and its antiparticle: they annihilate each other when they are brought together, or they can be produced together out of nothing (a “quantum fluctuation of the vacuum”). The interpretation in terms of logic is the law of contradiction: $a$ and its negation cannot both be true.

**Competition and cooperation**

A practical use of negative resources in reaction networks is that they offer us a simple way to express contradiction, inhibition or conflict, where the presence of a resource $a$ suppresses the presence of another resource $b$: $a \rightarrow -b$, which is equivalent to: $a + b \rightarrow \emptyset$ ($a$ and $b$ annihilate each other). This means that the more $a$ is produced, the more $b$ is consumed, and therefore the lower the concentration of $b$ becomes. The causal relation between $a$ and $b$ is then negative: increase in $a$ implies decrease in $b$, decrease in $a$ implies increase in $b$. More generally the relation “$a$ inhibits $b$” holds when:

$$\exists r \text{ such that } a, b \in I(r), b \notin O(r)$$

This means that $a$ is necessary to run a reaction $r$ that consumes, but does not produce, $b$. The opposite relation, “$a$ promotes $b$” applies when $a$ is necessary for a reaction that produces, but does not consume, $b$. The relations of inhibition and promotion can be (but need not be) symmetric, in which case $a$ and $b$ inhibit or promote each other. In this case, we might say that $a$ and $b$ are competitors, respectively cooperators.

Inhibition is a negative causal influence, promotion is a positive one. If you connect an uneven number of negative influences in a cycle, you get a negative feedback loop: the resources will indirectly suppress themselves. If you connect only positive influences, or include an even number of negative influences, the cycle becomes a positive feedback loop: the resources in the cycle will indirectly promote their own growth. A negative feedback loop leads to either an oscillation or a stabilization of the concentrations of the resources around an equilibrium level. Positive feedback produces an exponential growth of the resources. Systems dynamics (Sterman, 2000) is a very useful formalism for modeling networks of positive and negative causal influences, and the positive and
negative feedback loops they form. Compared to COT, however, it lacks the ability to combine different resources in one reaction.

Inhibition seems useful for modeling cognitive processes—both those expressed by formal logic, and those modelled by neural networks with inhibitory connections. A concrete application may be the modeling of competitive social communications in which one type of message (say x) denies, contradicts or suppresses another type (y).

Thus, reaction networks provide an elegant model of competitive or conflictual relations. The interactions between reactions take place through their shared resources (“nodes” or “vertices” in traditional network terminology): the same resource can appear in input and output sets of different reactions. Because these resources are either consumed or produced by the reactions, different reactions can either facilitate each other (e.g. when the one produces a resource needed by the other) or hinder each other (when the one consumes a resource needed by the other). This creates a complex, non-linear dynamic of cooperation (mutual amplification or synergy) and competition (mutual inhibition, conflict, or friction), which is the hallmark of self-organization.

Since the type and quantity of available resources will change under the effect of the reactions that are processing them in the medium they share, reactions too will be initiated, amplified, suppressed or shut down while the overall process is running. The general logic of self-organization, co-evolution and natural selection (Heylighen, 1992, 1999, 2008) will ensure that this process moves towards a self-sustaining regime, as resources and reactions that do not fit in with the emerging system are eventually eliminated, while those that efficiently exploit the more abundant resources grow and take over. This is the origin of the “organizations” that we will define further.

Organizations

The most important new concept introduced by COT is an organization. This denotes a reaction system that is fundamentally self-sustaining: the resources it consumes are also the resources it produces, and vice-versa. This means that although the system is intrinsically dynamic or process-based, constantly creating or destroying its own components, the complete set of its components (resources) remains invariant, because what disappears in one reaction is recreated by another one, while no qualitatively new components are added.

This property is an essential part of what is called autopoiesis or self-production, a concept that Maturana and Varela introduced to characterize living organisms (Maturana & Varela, 1980; Varela, 1979). Organizations are simpler than living systems, though, because they do not produce a spatial boundary separating them from their environment—the second fundamental characteristic of autopoietic systems. As such, they were introduced as simple models for the origin of life out of cycles of chemical reactions, before the emergence of the first cells that would enclose such autocatalytic cycles in a membrane, thus separating them from the environment.

Consider a subnetwork <M’, R> of a larger reaction network <M, R>, i.e. M’ ⊆ M. The formal definition of an organization is derived from three characteristics that such a reaction network <M’, R> can have:
• **closure**: this means that nothing new is generated: the only resources produced by the reactions are those that were already in the starting set \( M' \): \( \forall r \in R \text{ such that } I(r) \subseteq M' \), the requirement holds that \( O(r) \subseteq M' \).

• **semi-self-maintenance**: this is the complementary condition that nothing existing is removed; each resource consumed by some reaction is produced again by some other reaction working on the same starting set: \( \forall x \in M \text{ for which } \exists r \in R \text{ such that } x \in I(r), \exists r' \in R \text{ such that } I(r') \subseteq M, \text{ and } x \in O(r') \).

• **self-maintenance**: this is a stronger form of the semi-self-maintenance condition, which states that each consumed resource \( x \in M' \) is not only produced by some other reaction in \( <M', R> \), but that the amount produced is at least as large as the amount consumed.

The determination of self-maintenance is more complex than the other two conditions, because it requires the introduction of a quantitative dynamics in the reaction network, which specifies the rate at which resources are consumed and produced by the different reactions. This is necessary to establish the long-term maintenance of the resource set \( M \), because the reaction producing \( x \) may be slower than the one consuming it, so that the concentration of \( x \) eventually goes to zero. The speed or rate of production for each of the reactions is expressed by the “flux vector”. This then needs to be multiplied by the “stoichiometric matrix”, specifying which reactions consume and which produce a particular resource, in order to calculate the net balance of consumption/production. For simplicity, we will not discuss this quantitative aspect in the present qualitative description. Therefore, we will ignore the flux vector and the more complicated calculations that need to be performed on it in order to determine whether there exists a self-maintaining set of resources.

We are now able to define the crucial concept of **organization**: a subset of resources and reactions \( <M', R> \) is an organization when it is closed and self-maintaining. This basically means that while the reactions in \( R \) are processing the resources in \( M' \), they leave the set \( M' \) invariant: no resources are added (closure) and no resources are removed (self-maintenance).

This may seem a rather uninteresting property: nothing really changes. Most theories, models and formalisms are based on invariant elements, sets and structures, so what is novel here? The fundamental contrast with classical modeling frameworks is that we started by assuming that everything changes: all resources are in a constant flux, being consumed by some reactions, produced by others, but by default processed into something different. The concept of organization establishes that stability can arise even within this turbulent network of changes.

An organization is an emergent, self-producing system, which sustains itself only by reprocessing its components, and thus constantly rebuilding itself. This is the peculiar property of living systems that Maturana and Varela have tried to capture with their concept of **autopoiesis**. What COT adds is that the same kind of emergent organization can arise in a wide variety of other domains outside of biology, on the sole condition that we have a sufficiently rich network of reactions and resources. Moreover, COT reformulates the rather difficult and confusing notion of autopoiesis as a simple mathematical property characterizing even simpler sets of resources and reactions.
Some examples

The simplest organization would consist of the single resource a, and the single reaction: \( a \rightarrow a \). This would be the description of a resource that just maintains itself, without interacting with anything else. The organization becomes slightly more interesting when we add the reaction \( \rightarrow a \). Here a is not just maintained, it is also created out of nothing. We can make it more interesting by adding: \( a \rightarrow \). This means that a is not only produced or added, it is also removed from the medium (the “reaction vessel”). This would describe a situation where some resource fluctuates in and out of existence.

For the simplest non-trivial organization, we need two resources \{a, b\} that interact. They define an organization when the reactions form a cycle: \( a \rightarrow b, b \rightarrow a \). This can be extended with an unlimited number of intermediate stages:

\[
\begin{align*}
  a & \rightarrow b, \\
  b & \rightarrow c, \\
  c & \rightarrow d, \ldots, \\
  z & \rightarrow a.
\end{align*}
\]

This is still too simple to be very useful, but we can make it more complex by considering reactions with more than one input or output, e.g.

\[
\begin{align*}
  a + b & \rightarrow c, \\
  c & \rightarrow d + e + f \\
  e & \rightarrow a, \\
  d + f & \rightarrow b.
\end{align*}
\]

Here and a and a b together are transformed into a c, which is then converted to d, e and f, which again produce a and b, so that the cycle can start again. Let us make it more concrete by considering recognizable resources and reactions, in this case describing the organization of the Earth’s ecosystem at an abstract level.

\rightarrow \text{sunlight}

\begin{align*}
  \text{plants + sunlight + carbon dioxide + minerals} & \rightarrow \text{plants + oxygen} \\
  \text{plants + animals + oxygen} & \rightarrow \text{animals + carbon dioxide + detritus} \\
  \text{detritus + bacteria} & \rightarrow \text{bacteria + carbon dioxide + minerals}
\end{align*}

This describes the recycling of oxygen, carbon dioxide and minerals by plants, animals and bacteria, fueled by the energy of the sun (which enters the system from the outside, which is why the reaction producing it has no input within the system). This is subtler than a simple cycle, because reactions require several inputs while producing several outputs that are all needed to sustain the organization. But the system is properly autopoietic, as it produces all its essential components: nothing that is needed to sustain the organization gets lost; nothing new is added.

Note that some resources (such as bacteria in the last reaction) appear in both the input and output of a given reaction. That means that they are neither removed nor added by that reaction. Yet, they are necessary for the reaction to happen. In chemistry, such resources are called catalysts: they enable or facilitate a reaction, but are not themselves affected by it. In our more general
interpretation, we may call them agents (Heylighen, 2011): they act on the other resources in the reactions, processing them into something else. For example, the bacteria are the agents that turn detritus into the carbon dioxide and minerals that are needed by the plants. The plants are the agents that transform these resources, with the help of sunlight, into oxygen (and more plants). The animals act on the plants and oxygen, converting them to detritus and carbon dioxide, which then again function as “food” for respectively the bacteria and the plants.

**Extending the model**

This model of global recycling is of course much too simplified. To start with, it does not specify the relative proportions of the different resources produced and consumed. For example, plants do not produce just oxygen, they grow, thus producing more plants. In the multiset version of COT, the additional amount could be specified e.g. by writing “2 plants + oxygen” on the output side of the reaction. While this may clarify the relative proportions, the actual rate of production would need the full, quantitative version of COT, which includes the relative speeds of the different reactions as expressed by the flux vector. We will ignore these complications in the present introductory survey, and continue focusing on the power of COT for qualitative modeling.

Qualitatively, the simple model could be extended by noting some additional processes, such as plants → detritus (plants die, thus producing matter to feed bacteria), and animals → detritus (animals similarly die). We may also want to specify that it is not only bacteria that break down organic matter, but fungi as well, thus adding: fungi + detritus → fungi + carbon dioxide + minerals. But fungi are sometimes eaten by animals: fungi + animals → animals + carbon dioxide + detritus.

A different kind of extension may occur by making the general resource categories more specific. For example, we could note that not all animals eat plants or fungi, but that some are carnivores. This leads up to split up the category “animals” into the categories “carnivores”, “omnivores” and “herbivores”, each characterized by its own specific reactions. In this way, we can go on adding reactions and the concomitant resources until we feel the model is detailed enough to include everything that seems relevant for a realist description.

But the crucial question remains: is the resulting network an organization? By adding a particular reaction, we may create a “source” or a “sink” for a particular resource, either injecting it into a system in which it was previously absent (thus interrupting closure), or removing it from the system (thus interrupting self-maintenance). Let us then try to better understand how organizations emerge.

**Self-organization**

An arbitrary subset of a reaction network will in general not be an organization: its reactions working on its resources will produce additional resources (non-closure). These additional resources may react with some already present resources producing even further new resources. Thus, every addition may activate reactions that produce further additions. However, this process of growth of the resource base must come to an end when there are no further resources that can be produced by reactions working on the already present reactions. At that stage, all produced resources are already
in the present set, and closure is reached. Thus, closure can be seen as an attractor of the dynamics defined by resource addition: it is the end point of the evolution, where further evolution stops.

Let us now apply the same reasoning for self-maintenance, starting from the previously reached closed set. Some of the resources present in that set will be consumed by the reactions, but not produced, or at least not produced in sufficient amounts to replace the amounts consumed. These resources will therefore disappear from the closed set. Note that this does not affect the closure, because loss of resources cannot add new resources. Without these resources, some of the reactions producing other resources will no longer be able to run. Therefore, the resources they otherwise produce will no longer be replaced if they are consumed by some other reaction. If no other reactions continue producing these resources, they too will disappear from the resource set, possibly triggering the disappearance of even further resources that depend on them for their production. Thus resources disappear one-by-one from the set. However, this process too must come to an end, when the remaining resources do not depend for their production on resources that have been removed, but only on resources that are still being produced in sufficient amounts. Thus, self-maintenance too can be seen as an attractor of the dynamics defined by resource removal.

The combination of resource addition ending in closure followed by resource removal ending in self-maintenance produces an invariant set of resources and reactions. This unchanging reaction network is by definition an organization.

This scenario for the spontaneous emergence of an organization illustrates the general principle of self-organization (Ashby, 1962; Heylighen, 2001, 2008): any dynamic system will eventually end up in an attractor (originally called “equilibrium” by Ashby), i.e. a stationary regime of activity that the system cannot leave. In the attractor regime, the different components of the system have mutually adapted, in the sense that the one no longer threatens to extinguish the other: they have co-evolved to a “symbiotic” state, where they either peacefully live next to each other, or actively help one another to be produced, thus sustaining their overall interaction.

This is the default state of an evolved ecosystem—such as a forest or a coral reef—in which all the different species of plants and animals have adapted to the network of dependencies they all together constitute. While some of these species are predators of other species, they will normally not consume more of their prey than what is produced from other resources in the ecosystem. Their predation may actually control the population numbers of their prey to such a degree that the prey cannot increase in population so much that they exhaust other species on which they depend, thus indirectly threatening their own survival. A classic example are the wolves that were reintroduced as top predators to the Yellowstone natural reserve: their presence reduced the number of deer, thus allowing vegetation that the deer were consuming to recover, which in turn helped a wide variety of other species dependent on that vegetation to increase in number.

Note that the quantity of resources in the system is likely to fluctuate over time—e.g. following the classic predator-prey dynamics that leads to periodic increases and decreases, or following a more chaotic dynamics. But on the qualitative level, each of the resources should be able to be produced at a rate sufficient for it not to disappear altogether, because this would entail a potentially radical reorganization of the ecosystem.

The relations between the different resources and reactions in an organization form a similar web of predation, competition and cooperation—or, more generally, mutual dependencies—that has
stabilized into a self-sustaining network. Depending on the number and type of reactions, this network can be very complex. However, its defining features of closure and self-maintenance are easy to formulate mathematically, and to verify computationally—by running the reactions and checking whether each resource is produced at least as much as it is consumed, while no new resources are created.

**Sustainability and resilience**

An organization is by definition a self-sustaining, and therefore sustainable, system. That means that it can maintain perpetually, without ever running out of the resources that it needs to function—either because all resources are recycled through the inherent reactions, or because there is a dependable input from outside the system. Many organizations do not just maintain: they grow, because they produce more of certain resources than they consume (e.g. through a positive feedback cycle). Such resources are said in COT to be “overproduced” (Veloz, Reynaert, Rojas, & Dittrich, 2011). Such organizations fulfill the ideal of sustainable development: growth that can be sustained indefinitely.

Ecosystems are normally sustainable with an approximately constant level of resources. Economic systems, however, although they grow, are often unsustainable: they consume more of certain resources than they produce. Therefore, they are likely to collapse when the resource reserve is eventually exhausted. For example, our present economy is largely relying for its energy on oil and other fossil fuels that are in limited supply and cannot be renewed. Creating a sustainable economy means shifting to energy sources that are renewable, either through a dependable external input (e.g. wind energy) or through reprocessing within the system (e.g. biofuel derived from plants grown and harvested by the system).

In this example, the cause of unsustainability is easy to identify as it resides in a single resource (fossil fuel), and therefore the solution is obvious (replacing this resource by other, renewable resources). In general, sustainability emerges from the reactions between all the resources used, where a shortfall in one resource may be compensated by an increased production of another resource playing a similar function. It is here that we need the more sophisticated mathematical formalism of COT in order to establish in how far this reaction network is self-maintaining.

Related to the notion of sustainability is the one of resilience. Sustainability denotes the ability of the system to maintain and grow on its own, without outside interference. Resilience broadens this notion to the ability to maintain the essential organization even in the face of serious outside disturbances. A resilient organization is one that will survive and recover from shocks induced by the environment. In contrast, a fragile organization is one that is likely to disintegrate as soon as it encounters a serious disturbance.

In COT, a disturbance can be represented as the introduction of a new resource that reacts with some of the existing resources, thus interfering with the network of reactions that defines the organization. Another type of disturbance would be the disappearance of a resource that the organization relies on. In practice, both types of disturbances will reduce the availability of certain resources that are part of the organization, either by removing them at the input stage, or by inhibiting them via internal reactions. To survive such a disturbance, a resilient organization will
need to either suppress the source of the disturbance before it interferes with the organization’s critical “metabolism”, or to replace the lost resources before their absence makes further self-maintenance impossible. In other words, the organization will need to counteract or compensate the disturbance and/or its effects on the network of reactions so as to minimize the deviation from the viable configuration. This defines the cybernetic process of regulation or control (Heylighen & Joslyn, 2003).

The simplest method of control is buffering: maintaining a large enough reserve of resources so that temporary reductions in availability have little effect. This can be achieved by organizations that increase their resource base through overproduction of the most crucial resources. The next method is negative feedback: organizing the network of reactions in such a way that deviations from the desired concentration of resources are automatically counteracted after each cycle of consumption and production. For example, a reduced supply of a particular resource may automatically trigger an increased production of that supply by some other reaction. This kind of dynamics is common in metabolic pathways and in ecosystem interactions. For example, if foxes eat more rabbits, less rabbits will be left, and therefore some of the foxes will starve. A reduction in the number of predators will then let the rabbit population recover.

The third basic control method is feedforward: neutralizing the disturbance before it has had the chance to perturb the functioning of the system. This can be achieved by reactions that consume the disturbing resource before it could have interfered with other, vital resources. The tricky part here is that these neutralizing reactions can only be active when a disturbance is present for them to react with. This means that that most of the time these reactions remain “dormant”: the organization has the potential to react, but will only do so when the right condition is present. One way to implement such capability is by means of a collection of resources that are either overproduced, or are not consumed by any of the reactions in the organization, and so remain in reserve. Each of these resources can react with a particular disturbance, either getting consumed in the process, or, preferably, functioning as a catalyst that remains in the system after the reaction. The larger the variety of such potential “neutralizers” contained in the organization, the larger the variety of disturbances it can survive. This implements Ashby’s law of requisite variety (Ashby, 1958; Heylighen & Joslyn, 2003).

An example of such a collection of neutralizers are the genes of an organism that are activated via a particular chemical pathway whenever the cell encounters a particular disturbance. Once activated, these genes produce enzymes that neutralize the disturbance. But as long as a specific disturbance does not occur, the genes remain non-active snippets of DNA. Other examples of “dormant neutralizers” are antibodies, which are produced by the immune system in large quantities only in case of infection, and the armed forces of a country, which are mobilized only if the country is attacked.

We have argued that arbitrary networks of reactions will self-organize to produce sustainable organizations, for the simple reason that an organization is an attractor of their dynamics. It is less obvious that these organizations would also be resilient. However, evolutionary reasoning shows that resilient outcomes are more likely in the long run than fragile ones. First, any evolutionary process starts from some arbitrary point in the state space of the system, while eventually reaching some attractor region within that space. Attractors are surrounded by basins, from which all states lead into
the attractor (Heylighen, 2001). The larger the basin of an attractor, the larger the probability that the starting point is in that basin. Therefore, the system is more likely to end up in an attractor with a large basin than in one with a small basin. The larger the basin, the smaller the probability that a disturbance pushing the system out of its attractor would also push it out of the basin, and therefore the more resilient the organization corresponding to the attractor. Large basins normally represent stable systems characterized by negative feedback, since the deviation from the attractor is automatically counteracted by the descent back into the attractor.

The higher probability of starting from a large basin does not exclude the possibility of ending up in an unstable attractor, characterized by a small (or empty) basin. However, these unstable attractors will normally not survive long, as nearly any perturbation will push the system out of that attractor’s basin into the basin of a different attractor. After a number of such attractor-to-attractor shifts the probability increases that the eventual attractor will have a large basin, and therefore be stable. This very general, abstract reasoning makes it plausible that systems that are regularly perturbed will eventually settle down in a stable, resilient organization. This is an application of the “order from noise” principle, according to which increased variation (“noise”) facilitates the self-organization of a stable configuration (Heylighen, 2001; Von Foerster, 1960).

The shifting from one attractor to a similar, neighboring one can be seen as a higher process of evolution, in which the system adapts to changing conditions by changing its organization, but in such a way as to maintain maximal continuity. This means that subsequent organizations maintain most of their resources, while changing only a few. Such a “sideward shift” of organization normally happens as a combination of two “vertical” shifts, one “upward” that adds resources, and one “downward” that removes resources (Matsumaru, di Fenizio, Centler, & Dittrich, 2006). The ability of the system to undergo such minimal shifts in response to great disturbances exemplifies a higher level of resilience that may be called “evolvability”.

What is as yet unclear is how such organizations are precisely organized: what kind of resources, reactions and combinations between these reactions make up a resilient whole? A first theoretical decomposition of organizations (Veloz et al., 2011) gives us some hints about subsets of resources that play different roles—which may or may not contribute to enhanced resilience. A more detailed answer is likely to come from simulations, in which reactions systems are randomly generated, allowed to settle into an attractor, which are then repeatedly perturbed by generating random disturbances. We can then compare the organizations that were easily destroyed by disturbances with those that proved to be particularly resilient.

Another approach to the problem is the study of the metabolic networks used by real organisms. These appear to be surprisingly resilient in the face of random mutations removing or adding gene-regulated reactions (Matias Rodrigues & Wagner, 2009). A likely reason is the redundancy of pathways for producing or consuming critical resources, so that the loss of a pathway through mutation is simply compensated by more activity in other pathways. This is one of the factors that explain the great evolvability of living systems: they can afford to undergo a lot of variation without losing their essential ability to self-maintain. This allows them to explore an immense space of possible organizations and thus to discover ever more adapted ones.
Metasystem transitions and topological structures

One of the most important but as yet poorly understood processes in evolution is the *metasystem transition* (Heylighen, 1995; V. Turchin, 1995) or major evolutionary transition (Maynard Smith & Szathmáry, 1997): the emergence of a higher-order organization that is constituted out of lower-order organizations. Examples are the origin of living cells from autocatalytic cycles of reactions, the emergence of multicellular organisms out of single cells, and the emergence of societies out of individuals (Heylighen & Campbell, 1995). At first sight, COT does not allow us to describe such a process, as it recognizes only two levels of systems: individual resources and the organizations constituted out of such resources. However, an organization S behaves in a way like a resource of the catalyst type: it is invariant under reactions, yet it still has in general an input of resources I(S) it consumes, and an output of resources O(S) it produces. This allows us to summarize the activity of the organization by the following “higher-order” reaction:

\[ I(S) + S \rightarrow S + O(S) \]

Suppose that \( I(S) = \{a, b\} \) and \( O(S) = \{c, d, e\} \), then we can write this as a more conventional reaction:

\[ a + b + S \rightarrow S + c + d + e \]

The fact that S is itself constituted of a network of resources and reactions does not really make any difference when seen from the outside. S behaves like a “black box” which processes a given input into a specific output. If S is a resilient organization, it can maintain itself even when the input changes, producing a correspondingly changed output of “waste products”. This means that S behaves like a multipurpose catalyst, or “agent”, which converts certain inputs into certain outputs.

This means that it is in principle possible to describe a *meta-organization*, as a closed and self-maintaining network of resources and reactions that contains several organizations like S among their catalytic resources. In practice, the problem is to maintain a separation or distinction between these suborganizations. Since all resources are freely shared within the medium, it is likely that an organization S will produce or consume some of the resources used by another organization S’. This means that the two organizations will interfere with each other’s reactions. This will trigger a process of mutual adaptation whose end result is likely to be a single, merged organization, including reactions from both S and S’. But if all organizations within a meta-organization eventually merge, then there is no longer any sense in distinguishing between the organization level and the meta-organization level: everything just settles down into a single organization without distinguishable components.

In concrete systems, such as organisms or ecosystems, this problem is avoided because of the existence of topological boundaries: the resources within one cell or organism do not mix with the resources in another cell or organism, because they are kept apart by some kind of membrane or spatial separation. This suggests a straightforward extension to the COT formalism: introduce a topological structure that subdivides the medium into “cells” separated by membranes or boundaries.
across which resources can be exchanged in a restricted manner between neighboring cells (Peter et al., 2011). This is a classic modeling approach used in simulations of complex adaptive systems (Holland, 2012; Miller & Page, 2007). For simplicity, such simulations typically situate their agents in the cells of a two-dimensional, rectangular grid rather than in a more realistic three-dimensional, continuous space. However, this approach seems rather artificial, because the spatial structuring of organizations is imposed externally and rigidly, instead of being the product of internal self-organization. Therefore, we will explore in how far topological structure may emerge from the abstract network of resources and reactions that constitutes a reaction system.

There is a simple “trick” to separate resources into distinct cells: label each resource with the number of the cell in which it is supposed to be situated (Peter et al., 2011). For example, a1, a2 and a3 can be seen as instances of the same resource a, but localized in respectively cell 1, cell 2 and cell 3. This merely produces a partitioning of resources. To introduce a topology, we moreover need to specify which cell is a neighbor of which other cell. For example, if the three cells are arranged on a line, then cell 1 neighbors cell 2, and cell 2 neighbors cell 3, but cell 1 does not neighbor cell 3. The neighboring relation can be expressed by a reaction that describes the movement or diffusion of a resource from one cell to its neighbor, e.g.

\[
\begin{align*}
a_1 &\rightarrow a_2 \\
a_2 &\rightarrow a_3
\end{align*}
\]

This means that resource a can move from cell 1 to cell 2, and from 2 to 3, but not directly from cell 1 to cell 3. Using this method to specify neighboring relationships, we can impose a variety of topological structures on the collection of cells, without need to remain stuck with rectangular grids.

Still, this method remains artificial in the sense that we started by defining cells, and then labeled resources according to this imposed topology. In a basic reaction system there are only resources and reactions, no cells or labels. Let us therefore try to reconstruct a topology if we ignore the labels and just consider differently labeled resources (say a1, a2, a3, …) as though they were a priori different resources (say x, y, z, …). The “neighboring” reactions would then become:

\[
\begin{align*}
x &\rightarrow y \\
y &\rightarrow z
\end{align*}
\]

In a typical reaction, a single resource (x) does not spontaneously turn into a different resource (y) without any other resource being consumed or produced. Therefore, we can reinterpret such one-to-one reactions as cell-to-cell transitions: only the cell changes, the intrinsic qualities of the resource (which other resources it can react with) remain the same. That leaves us with all the many-to-one, one-to-many, and many-to-many reactions. Let us call the set of these remaining reactions \(R_{\text{many}}\). Resources participating in a reaction from \(R_{\text{many}}\) may be assumed to be present in the same cell: two resources can only react with each other if they are in direct, local contact, i.e. in the same cell. Moreover, their products can be assumed to be formed within this same cell. Therefore, two resources that participate in the same reaction \(r \in R_{\text{many}}\) must be present in the same cell.
This relationship of co-presence is reflexive (x co-present with x) and symmetric (if x co-present with u, then u co-present with x). Co-presence within the same cell would also imply that the relationship is transitive (if x co-present with u, and u with v, then x co-present with v), so we need to construct the transitive closure of the relationship defined above. This makes co-presence into an equivalence relation that will automatically partition the set of resources into subsets corresponding to different cells. Thus, we can induce a partition of the resources simply by examining which resources can react with which other resources. This means that resources present in different cells by definition cannot react with each other. Therefore, the organizations using these reactions cannot interfere, and must remain separate. This solves the problem of having autonomous organizations functioning within the same reactive network.

The processes of diffusion between the cells represented by the one-to-one reactions linking different cells then induce a topological structure by specifying which cells are neighbors (and which resources may or may not cross the “membrane” separating them). Through diffusion, the output of one cell can become the input of another cell. This enables the emergence of a meta-organization that spans several interacting cells. Thus, we have formulated a foundation for a mathematical model of the emergence of a topological structure that both separates and connects local, cell-based organizations.

Note that the cells in this construction are independent of physical space. Thus, it is in principle possible to have two separate organizations occupying the same space, while interacting only indirectly. An example could be a linguistic and an economic system co-existing in the same human society. The resources of the linguistic system would be words, and those of the economic system would be goods. Words can “react” with each other by forming sentences, while goods can react by being used or exchanged simultaneously. Words and good interact indirectly in the sense that words may be produced in order to initiate an exchange of goods, while the receipt of the good may initiate the production of words. But a word and a good together do not create a new resource because they belong to different realms, dimensions or aspects of the human society. In systems theory, such systems that are distinct yet spatially inseparable have been called “aspect systems” (in contrast to subsystems, which are spatially separate).

This mathematical construction needs to be elaborated further, and explored by means of computer simulations of the self-organization of randomly generated reaction network. However, it suggests that it is in principle possible to introduce a topology and a mechanism of metasystem transition within COT without imposing additional constraints. This opens the door to the modeling of the multilevel self-organization and dynamical hierarchies (Rasmussen, Baas, Mayer, Nilsson, & Olesen, 2001) that are necessary for understanding the origin and evolution of life and society.

Concrete applications

Here we wish to give a brief survey of existing and future application domains for COT.

Most obviously, COT has been used to simulate networks of chemical reactions with the focus on the emergence of stable systems. One example is the modeling of the chemistry of a planetary atmosphere (Centler & Dittrich, 2007). The initial application that inspired the development of COT was to model how such chemical networks could develop the degree of
autonomy that we associate with simple living creatures (Fernando & Rowe, 2008). Previously, this problem of the origin of life had been approached by looking for autocatalytic cycles of chemical reactions (Hordijk, Hein, & Steel, 2010; Maynard Smith & Szathmáry, 1997; Vasas, Szathmáry, & Santos, 2010). These are a more limited kind of organization that are both more difficult to build by evolution, and less flexible and resilient than more general chemical organizations. A related application domain is the study of complex metabolic networks in existing organisms, such as the bacterium E. Coli (Centler, di Fenizio, Matsumaru, & Dittrich, 2007). This domain has recently attracted a lot of attention under the label of “systems biology”, but still lacks an integrated theoretic framework (Machado et al., 2011). Another classic problem within this domain is the modeling of gene regulatory networks, in which genes activate or deactivate each other, resulting in a variety of regulatory “organizations” characterized by stable subsets of activated and deactivated genes. These would be the attractors of a complex network of reactions having the following form:

\[
\begin{align*}
\text{active gene 1} & \rightarrow \text{active gene 1} + \text{protein 1} & \text{(protein expression of an active gene)} \\
\text{active gene 2} + \text{protein 1} & \rightarrow \text{non-active gene 2} & \text{(expressed protein deactivates gene)} \\
\text{non-active gene 3} + \text{protein 2} & \rightarrow \text{active gene 3} & \text{(expressed protein activates gene)}
\end{align*}
\]

Because the reactions defining COT are intrinsically abstract, computable processes, they can be used as a foundation for a new method of computation, based on “artificial chemistries” (Matsumaru, Centler, di Fenizio, & Dittrich, 2005). Such “chemical computation” can for instance be used to program distributed artefacts (Matsumaru, Hinze, & Dittrich, 2009), helping them to coordinate their actions.

Once we make abstraction of the molecules that originally inspired COT, the application domain immediately extends to the social sciences, where the resources to be produced by reactions can e.g. be economic goods (Dittrich & Winter, 2005) or political decisions (Dittrich & Winter, 2008). In the latter case, the self-sustaining network of decisions producing further decision provides a simple formal model of the notoriously difficult theory of autopoietic social systems developed by Niklas Luhmann (Luhmann, 1986, 1995). Another application in the social domain is the modelling of evolutionary games through which individuals may or may not come to cooperate (Veloz, Razeto-Barry, Dittrich, & Fajardo, 2014).

We have alluded several times at the as yet unexplored potential for applying COT to problems in ecology, sustainable development, and the resilience of social and ecological systems. A related issue is the understanding of business ecosystems (Moore, 1996; Nachira, Nicolai, Dini, Le Louarn, & Leon, 2007; Peltoniemi, 2006), where different companies producing and consuming different goods and services form a symbiotic, co-evolving network where the one provides the resources for the others. Existing models in this domain, such as food webs or systems dynamics, tend to be limited to networks of one-to-one interactions, in which one variable (e.g. a predator population) positively or negatively affects one other variable (e.g. a prey population). In COT, we can examine how several resources in combination produce or consume a combination of other resources. While at first sight this makes modelling more complicated, the mathematics of COT shows that such multi-resource reactions makes the emergence of stable organizations easier rather than more difficult.
Another advantage of COT is that you can freely mix resources of very different types, such as organisms, chemicals, economic goods, and even human decisions. This makes it eminently suitable for modelling the truly complex social-technological-economical-ecological-physical systems that surround us, such as cities, businesses, regions, or our planetary society. This is the objective of the new approach of global systems science (Helbing, 2013; Helbing, Bishop, Conte, Lukowicz, & McCarthy, 2012).

Of course, modelling systems with that many variables cannot be done manually, but COT lends itself readily to the development of simple, modular computer programs that can examine a wide range of possible situations, and that are easy to extend or update. What still needs to be investigated is in how the algorithmic complexity of COT models grows as the number of resources and reactions increases.

To further illustrate the generality of COT, we wish to briefly suggest some more speculative applications. The section on resilience noted that an evolved organization is likely to exhibit a variety of regulatory mechanisms characteristic of a cybernetic system. Such a system is in essence a goal-directed agent, acting upon its environment in order to secure the resources it needs, while suppressing any disturbances it may encounter. That means that it exhibits not just the most basic features of life, but of cognition, intelligence and intentionality (Heylighen, 2011). Like all living systems, the goal or intention of an organization is to maintain and grow. To achieve this, it needs to produce the right actions for the right conditions (e.g. produce the right resource to neutralize a particular disturbance). This means that it implicitly contains a series of “condition-action rules” that play the role of the organization’s “knowledge” on how to act in its environment. The capability of selecting the right (sequence of) action(s) to solve a given problem constitutes the organization’s “intelligence”. To do this, it needs to perceive what is going on in its environment, i.e. to sense particular conditions (the presence or absence of certain resources) that are relevant to its goals. Thus, an organization can be seen as a rudimentary “intelligence” or “mind”.

Because this abstract conceptualization is independent of any concrete substrate, such as a brain, it is applicable to systems that exhibit intelligent behavior but that are otherwise very different from the individual humans that we tend to see as the sole possessors of minds. Examples are the intelligence exhibited by insect societies, plants, bacterial colonies, human organizations, the planetary ecosystem (“Gaia”), and the Internet functioning as a “Global Brain”. In all these cases, intelligence is distributed: it is not localized in some central decision-making component, but it emerges from the coordinated interactions between all components. Providing simple models of such distributed, dynamic organization is precisely the strength of COT.

Even the human brain is a complex, distributed network, where all the important features such as intelligence, intentionality and consciousness are emergent rather than localized in some specific neuron or assembly of neurons. Recently, great progress has been made in the understanding of consciousness as a pattern of activity that takes over the “global neuronal workspace” in the brain (S. Dehaene, Kerszberg, & Changeux, 1998; Stanislas Dehaene, 2014). For conscious processing of thoughts we need to maintain a pattern of activity long enough in our working memory so that it can be examined and processed by different brain modules. This is intrinsically difficult because neural activation cannot stay in the same place: a neuron that is excited by an electrical signal (“action potential”) cannot retain that electrical charge but must pass it on to one or more neighboring
neurons via its outgoing axon ending in synapses. If a sufficient number of incoming synapses pass on a signal, the newly reached neurons will get activated as well, passing on this activation via their outgoing synapses to further neurons. This transmission of activation can be described as a reaction of the form: \( a + b + \ldots \rightarrow e + f + \ldots \), where \( a \) and \( b \) are the initially activated neurons whose combined activation is necessary to activate the subsequent neurons \( e, f, \) etc. We may say that the activation of \( a \) and \( b \) is “consumed” by the reaction in order to “produce” the activation of \( e \) and \( f \).

What the neuronal workspace theory proposes is that “conscious” patterns of activation, in contrast to subconscious or subliminal processes, are to some degree self-sustaining: activation that leaves a neuron comes back to it at a later stage after having propagated through some complex, closed network. This creates coherent assemblies of neurons that are firing in a rhythmic, coordinated manner, so as to keep the idea “alive” long enough for it to be monitored and processed in a controlled, goal-directed manner—the hallmark of consciousness. Mathematical models of this process have been built (S. Dehaene et al., 1998), but they are quite complicated, making many ad hoc assumptions about specific neurophysiological properties and structures, while being able to simulate only the most basic dynamics of a neuronal assembly reaching “ignition” (self-sustaining activation). By interpreting coherent neuronal activation patterns as organizations, we may be able to reach a simpler, broader and more qualitative understanding of the different conscious patterns that the brain can produce, and how such patterns can shift into different patterns as new stimuli make them deviate from their initial organization, thus producing a “train of thought” or “stream of consciousness”.

References


