Autocatalytic Sets:
Complexity at the Interface of Chemistry and Biology

March 2022

Dr. Wim Hordijk
The Evolution Institute

JOHN TEMPLETON FOUNDATION
Inspiring Awe & Wonder

Table of Contents

I. Introduction

Life (biology) is more than the sum of its constituent molecules (chemistry). It depends crucially on the specific interactions between large numbers of molecules in the chemical reaction networks that give rise to living organisms. In particular, life depends on catalysts, i.e., molecules that speed up and regulate chemical reactions, with these catalysts being produced by the very same chemical reaction networks that they regulate. In other words, the reaction networks underlying living systems are self-regulating and self-sustaining.

One way to model and study such self-sustaining reaction networks is with the concept of autocatalytic sets. An *autocatalytic set* is a chemical reaction network in which the molecules mutually catalyze each other's formation from a basic food source. Autocatalytic sets have been studied extensively both theoretically and experimentally. For example, they have been shown to emerge spontaneously in various types of chemical networks, to be (potentially) evolvable, and to exist in the metabolic networks of actual living organisms. It has also been argued that the notion of autocatalytic sets can be applied to interaction networks beyond chemistry, such as ecosystems or even the economy.

Autocatalytic sets are a specific example of complex systems. A *complex system* is generally defined as consisting of many interacting parts, with the local interactions between these parts giving rise to a global, system-wide behavior that is difficult (or even impossible) to predict from the properties of the individual parts. In other words, the behavior of the system as a whole *emerges* from the actions and interactions of the parts, but cannot be directly derived from them. In an autocatalytic set, the chemical network as a whole is self-sustaining and evolvable, whereas (in general) none of the constituent molecules are. The network's self-sustainability and evolvability are dynamic, emergent properties, arising out of the many interactions (i.e., catalyzed chemical reactions) between the parts (i.e., molecules).

This research synthesis presents a general (and gentle) introduction to the concept of and research on autocatalytic sets, specifically from a complex systems perspective. It reaches across several traditional scientific disciplines, from mathematics, computer science, experimental chemistry, and evolutionary biology, to ecosystems and economics. And it involves multiple typical complex systems topics, such as interaction networks, dynamical systems, self-organization, emergence, and evolution. It thus provides a comprehensive overview of progress in complexity science based on a particular (and well-studied) example at the interface of chemistry and biology.

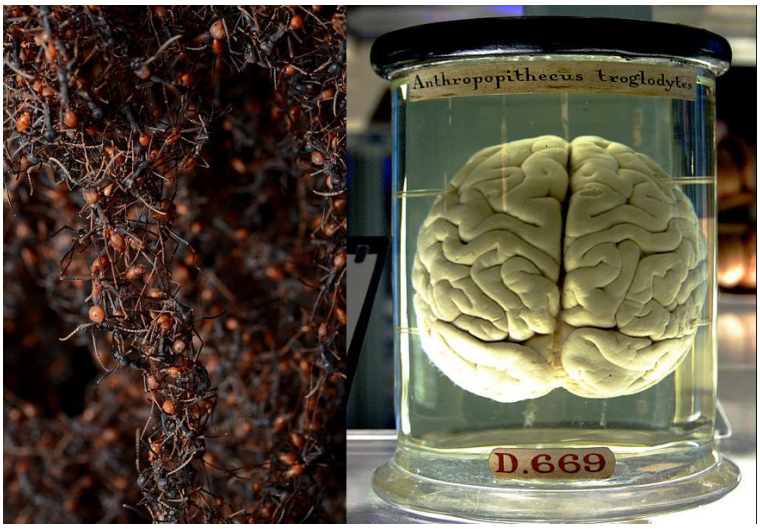
Back to Table of Contents

II. The Complexity of Life

Complex systems

Complex systems are commonly (and somewhat casually) described as systems where "the whole is more than the sum of its parts." Standard examples are an ant colony and the brain.

An individual ant follows a set of relatively simple rules that dictate its behavior. Yet put thousands of them together, and something bigger emerges out of the myriad interactions between them: a thriving colony that builds an intricate nest, has social structure in the form of division of labor, and manages to solve difficult tasks such as finding the shortest path between the nest and a food source. Moreover, all this emergent behavior happens without any central control, and arises purely out of the many local actions of and interactions between the individual ants.



Ant colonies and brains are standard examples of complex systems. Ant image: Geoff Gallice. Brain image: Gaetan Lee. CC BY 2.0

Similarly, your brain consists of thousands of neurons, where an individual neuron is basically a chemical switching device that gathers input signals from nearby neurons, and sends out signals itself based on these input signals. You would not call an individual neuron intelligent in any way. Yet hook many of them up in the right way, and intelligence (i.e., problem solving ability) emerges out of the local interactions (through chemical signals) between them.

Moreover, as with an ant colony, there is no central control in the system.

Although there is no single agreed upon formal definition (yet), there is general consensus on what constitutes a complex system. Computer scientist Melanie Mitchell, in her highly readable introductory book, *Complexity: A Guided Tour*, defines a complex system as "a system in which large networks of components with no central control and simple rules of operation give rise to complex collective behavior, sophisticated information processing, and adaptation via learning or evolution" (Mitchell, 2009; p. 13).

Physicist Stefan Thurner and co-authors, in their more technical book, *Introduction to the Theory of Complex Systems,* define complex systems as "systems whose states change as a result of interactions and whose interactions change concurrently as a result of states" (Thurner et al., 2018; p. v). They add: "Due to this chicken–egg-type problem, complex systems show an extremely rich spectrum of behavior: they

are adaptive and co-evolutionary; they show path-dependence, emergence, power laws; they have rich phase diagrams; they produce and destroy diversity; they are inherently prone to collapse; they are resilient, and so on" (Thurner et al., 2018; p. v).

Many biological systems (such as ant colonies and brains) are complex systems, whether they consist of many molecules interacting in specific ways to produce a living cell, many cells interacting to form a functioning organ or entire organism, or many organisms interacting to give rise to a diverse and thriving ecosystem. But complex systems also exist beyond biology, such as the economy, the internet, cities, and the healthcare system.

The science of complex systems aims to find the commonalities between all these different complex systems, to understand how their emergent properties arise out of the underlying local interactions, and (eventually) to control existing or design new complex systems. This science is highly interdisciplinary. It involves aspects from, among other fields, mathematics and physics (dynamical systems, networks, information theory), chemistry and biology (self-reproduction, evolution, ecosystems), economics and sociology.

The Santa Fe Institute (SFI) was the world's first research institute dedicated to the study of complex systems. It was founded in 1984 and is still one of the leading institutes for complex systems science, with a world-wide network of associated complexity researchers (including Mitchell and Thurner). In our increasingly complex society, complexity science plays an ever more important role. As the institute's mission statement says: "As we reveal the unseen mechanisms and processes that shape these evolving worlds, we seek to use this understanding to promote the well-being of humankind and of life on earth."

Life's organization

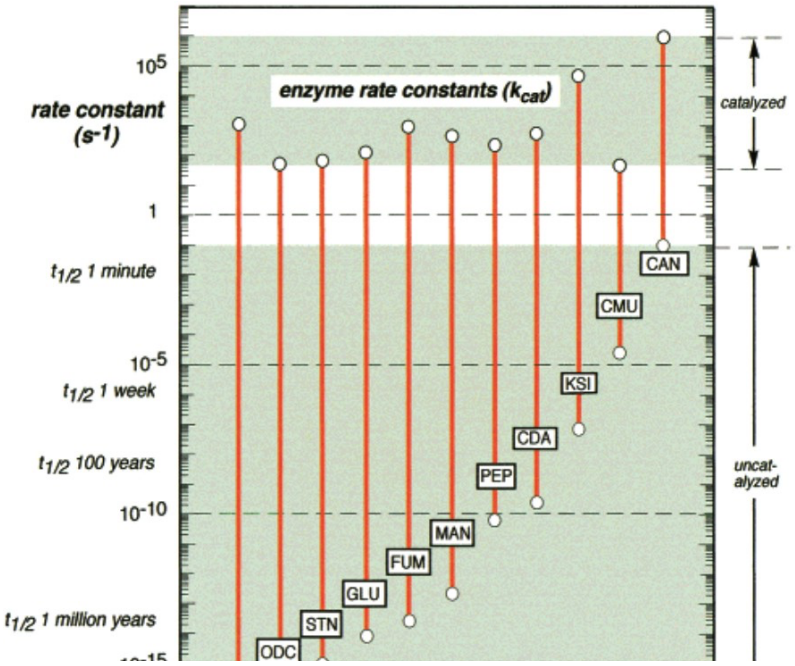
What is it that makes this "life on earth" so special, and so complex? To help answer this question, imagine the following experiment. Take some bacteria, for example, *Escherichia coli* (or *E. coli*, for short), and put them in a petri dish with appropriate nutrients. After just a couple of days the dish will be brimming with *E. coli* bacteria. Now take those same initial few bacteria, grind them up into their constituent molecules, put these in a petri dish with the same nutrients, and watch what happens. Nothing.

Next, consider an experiment that was done more than 50 years ago by biophysicists Arthur Skoultchi and Harold Morowitz (who was a longtime SFI associate). Take dried fertilized eggs from the common brine shrimp (Artemia), put them in liquid helium at a temperature of 2K (i.e., near absolute zero), and leave them for six days. Then slowly warm them up to room temperature, and watch what happens. Artemia will continue its normal life cycle: the eggs hatch, and the larvae will grow into adults that will mate and lay fertile eggs (Skoultchi & Morowitz, 1964).

> In both experiments life was destroyed, either by grinding or by freezing it to death. However, in the second experiment it was possible to restore the "living state." So what is the difference between these two experiments? The crucial distinction is that in the second experiment

the system's chemical organization was not destroyed.

Clearly, life is more than the sum of its constituent molecules. Indeed, as already stated above, living systems are complex systems. They have a particular chemical organization, i.e., they are based on chemical reaction networks in which the molecules interact and work together in such a way as to form a self-sustaining and evolvable (i.e., living) system. More precisely, living systems produce their own catalysts, which in turn maintain and regulate the very chemical reaction networks that produced these catalysts in the first place.



A comparison of reaction rates with (top) and without (bottom) enzyme catalysts for several organicreactions. The rate differences span many orders of magnitude. From Wolfenden & Snider (2001).

In chemistry, a catalyst is a molecule that speeds up (and regulates) the rate at which a reaction happens. However, a catalyst is not "used up" in that reaction. A single catalyst can catalyze multiple reactions over time. Life depends crucially on catalysts and could probably not exist without them, as the required chemical reactions would happen too slowly, or they would not be properly synchronized with each other.

In living systems, most catalysts are proteins, which are long chains of amino acids that fold up into complicated three-dimensional structures that determine their functionality (e.g., which reactions they can catalyze). Protein catalysts are generally referred to as enzymes. There is a little more chemical detail to enzymes, but we will return to that later on. For now it is important to know that most of these enzymes do not form spontaneously, but can only be produced by living systems. Hence life's intriguing circularity (or self-referentiality): life requires catalysts that can only be produced by life itself.

To capture this almost paradoxical self-referentiality of living systems more formally, in 1971 biologist Stuart Kauffman (long-time SFI associate and founding member) introduced the notion of autocatalytic sets.

Back to Table of Contents

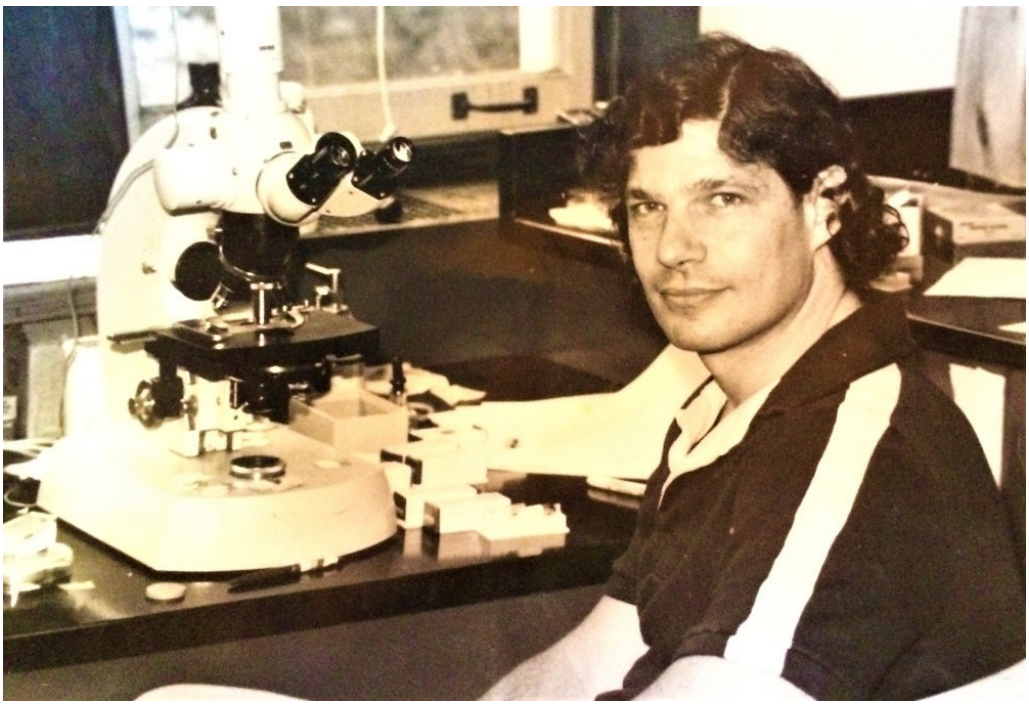
III. Autocatalytic Sets

Initial idea

The word *autocatalytic* literally means "self-catalyzing." In chemistry, an autocatalytic reaction refers to a chemical reaction where one of the products is able to catalyze the same reaction that produced it. Although there are examples of such autocatalytic reactions, they are not very common. However, this notion can also be generalized to a set (collection) of reactions.

Informally, an autocatalytic set is a chemical reaction network in which the molecules mutually catalyze each other's formation from smaller building blocks (the "food source"), such that the set as a whole is autocatalytic. As Kauffman explained: "Replication is the property of a complex dynamic system, not a single molecule. More fundamentally, self-replication is an autocatalytic process in which a set of molecules catalyzes the formation of a nearly identical second set. No molecule need catalyze its own formation" (Kauffman, 1971; p. 90).

He then goes on to argue what it would take to get such a set of collectively and autocatalytically reproducing molecules, focusing mostly on macromolecules such as peptides (short proteins). He even describes results from simple computer simulations (already back in 1971!), where different "polymers" (abstract peptides) have a certain probability of catalyzing various chemical reactions between these same molecules. In his own words: "Autocatalytic sets began to emerge when the probability of a molecule affecting a reaction was about 0.003 to 0.005" (Kauffman, 1971; pp. 94—95). He then ends with the following bold claim: "These global behaviors of macromolecular systems should underlie all organisms, no matter how evolution selected the surviving forms" (Kauffman, 1971; p. 95).



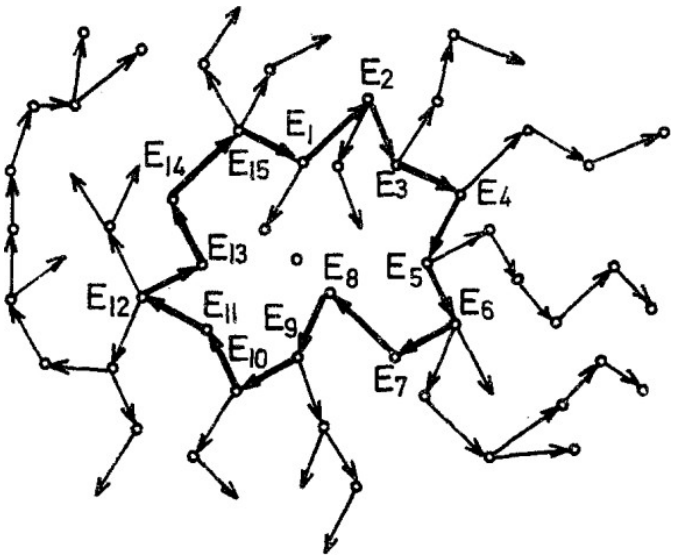
Stuart Kauffman in the lab at the University of Pennsylvania in the early 1970s. Image courtesy of Stuart Kauffman.

Since the molecules in the set mutually catalyze each other's formation, an autocatalytic set is self-sustaining (given the food source), or even self-reproducing. This is a property of the system as a whole that emerges out of the underlying reaction network, as none of the constituent molecules need be self-reproducing (as Kauffman already explicitly noted). Furthermore, as will be shown later, autocatalytic sets are also (in principle) evolvable. In other words, they can potentially adapt, diversify, and become more complex, which are again emergent (network) properties not inherent in the constituent molecules.

As Kauffman claims, autocatalytic sets are believed to be an essential underlying mechanism of life, allowing living systems (biology) to emerge from molecular reaction networks (chemistry). Thus, autocatalytic sets are a prime example of complex systems at the interface of chemistry and biology. Indeed, they fit the definitions of Mitchell and Thurner quite well. Autocatalytic sets are networks of components (molecules) with no central control and simple rules of operation (chemical reactions) giving rise to complex collective behavior (self-sustainability, self-reproduction) and adaptation via evolution. And as these autocatalytic networks evolve, they may concurrently change their interactions (catalyzed reactions) as a result of states (molecule types). For example, a newly arising variant of an existing catalyst may catalyze the same reaction more efficiently, or even become able to catalyze an entirely different reaction.

Despite Kauffman's initial results and claims, though, the idea of autocatalytic sets did not immediately attract much attention. This was partly because it was hidden in an appendix of a publication that was mostly about another concept that Kauffman had come up with just a few years earlier: random Boolean networks (an idea that did attract much attention right away). But perhaps even more so because of an influential paper that appeared in the very same year.

Also in 1971, Nobel laureate Manfred Eigen (former SFI science board member) published a paper in which he proposes almost exactly the same idea as Kauffman. Eigen explicitly considers reaction networks of proteins, where some of these proteins are able to catalyze the formation of others (indicated by arrows in the example on the right). If such a network contains a closed loop, then it forms what he calls a "catalytic network" (Eigen, 1971; p. 499). Eigen suggests that it may be quite plausible that such catalytic networks (i.e., closed loops) form spontaneously in large enough (random) sets of proteins. However, he then dismisses the idea again based on the claim that such catalytic networks would not be evolvable, and could thus never lead to real (complex) life.



Eigen's illustration of a closed loop (bold arrows)in a network of mutually catalyzing proteins. From Eigen (1971).

In other words, (auto)catalytic networks of proteins may have a high probability of forming, but they lack evolvability, which is an essential (and emergent) property of life. Eigen then goes on to introduce his idea of the hypercycle, which, at least in its original conception, combines nucleic acids (as information carriers, or genotypes) and proteins (as catalysts, or phenotypes). However, as will be shown later, autocatalytic sets are (in principle) evolvable, just not in the same genotype-phenotype way as considered by Eigen.

Either way, the idea of autocatalytic sets was not given much attention by anyone, except for Kauffman himself. Although not for another 15 years.

Early work

In 1986 Kauffman returned to his earlier idea of autocatalytic sets with the publication of a full paper on the topic. He states his main goal right away: "Catalytic 'closure' must be achieved and maintained. That is, it must be the case that every member of the autocatalytic set has at least one of the possible last steps in its formation catalyzed by some member of the set, and that connected sequences of catalyzed reactions lead from the maintained 'food set' to all members of the autocatalytic set" (Kauffman, 1986; pp. 2–3). He then develops a mathematical argument for his claim that such catalytic closure may be "highly probable." For this, he first presents a more detailed description of his earlier computer model.

Assume a set of abstract polymers up to (and including) length *M*, built up of two monomer types, say 0 and 1. (Kauffman used 'A' and 'B' in his initial work, but later on 0 and 1 were used, also by others; the results obviously remain the same.) These monomer types can be seen as the abstract equivalent of, for example, amino acids, which are the monomeric building blocks of proteins.

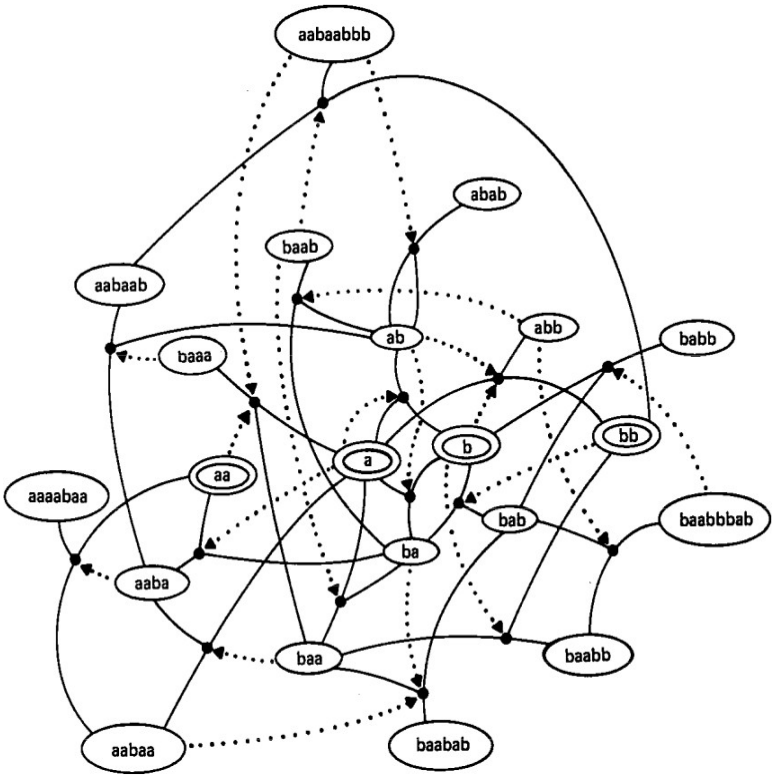
Next, assume there are two types of chemical reactions possible between these different polymers: (1) ligation of two polymers into a longer one, and (2) cleavage of a polymer into two shorter ones. An example of a ligation reaction is 000 + 11 --> 00011, and an example of a cleavage reaction is 010101 --> 0101 + 01. Note that the maximum length *M* is maintained, i.e., ligation reactions that produce polymers longer than length *M* are not allowed. Also note that for each ligation reaction, there is a corresponding (i.e, reverse) cleavage reaction.

Finally, assume that there is a probability *P* that a given polymer can catalyze a given reaction. In other words, for each possible pair of a polymer and a reaction, decide with probability *P* whether that polymer catalyzes that reaction. This probabilistic catalysis assignment is done once and then kept fixed, generating one particular instance of the model, i.e., a ligation/cleaveage reaction network with a particular assignment of which molecules catalyze which reactions. To create another instance of the model, these catalysis assignments are randomly assigned anew, again with probability *P*.

Kauffman then uses a well-known theoretical result on random networks by two Hungarian mathematicians (Erdős & Rényi, 1959; 1960) to show that autocatalytic sets will form almost certainly in instances of his abstract model when the diversity of polymers (i.e., the maximum polymer length *M*) is large enough given a (fixed) value for the probability of catalysis *P*. In other words, when keeping the parameter *P* fixed and slowly increasing the parameter *M*, there will be a "phase transition" (or "threshold") where suddenly autocatalytic sets start showing up in (random) instances of the model. This will be the case even for very small food sets consisting of, for example, only monomers and dimers (i.e., 0, 1, 00, 01, 10, and 11).

Note that Kauffman did not require these autocatalytic sets to consist of the entire reaction network, i.e., involving all polymer types up to length *M* and all possible reactions between them. In fact, an autocatalytic set can consist of only a subset of the full reaction network. However, as a consequence of the mathematical argument used by Kauffman, autocatalytic sets would be expected to consist of most (if not all) polymer types in the reaction network instances of his model.

Around the same time, Kauffman also teamed up with physicists Norman Packard and Doyne Farmer (also a long-time SFI associate). In their joint paper, they include an explicit example of an autocatalytic set that exists within an instance of the abstract polymer model with *M*=8.



An example of an autocatalytic set within an instance of Kauffman's model. The molecules are polymers (with monomers 'a' and 'b'), shown inside ovals. The food set is indicated with double ovals. The reactions are (bidirectional) ligation/cleavage, represented by black dots with solid lines going from thereaction to its reactants and products. Catalysis is represented by dotted arrows going from a catalyst to a reaction. From Farmer et al. (1986).

However, the existence of an autocatalytic set within the reaction network does not guarantee that it will actually show up (or be "realized") in the overall *dynamics* of the system. As the authors state: "Once chemical kinetics are taken into account competition for resources limits growth. Differences in efficiency can produce drastic differentials in concentration, so that many species are effectively not present in the system at all. Thus, even though it complicates matters considerably, a consideration of kinetics is essential for a realistic assessment of the potential to generate autocatalytic sets" (Farmer et al., 1986; p. 58).

In fact, it is a common property of complex systems that their dynamical behavior cannot be fully predicted in advance, other than explicitly "running" the system. In other words, even if it is possible to write down a set of mathematical equations that accurately describes the system's dynamical behavior, it is generally not possible to solve those equations and predict the state of the system at any given time in the future. In the case of chemical reaction networks, it is possible to write down a set of differential equations that describes accurately how the concentrations of the different molecule types change over time. However, due to theself-referential nature (i.e., catalytic closure) of autocatalytic sets, this set of differential equations can (in general) not be solved theoretically. They can only be simulated numerically, small time step by small time step, to know what the state of the system (i.e., the concentration of each molecule type) will be at any later time.

And this is exactly what Packard and Farmer did. For several instances of Kauffman's polymer model, they wrote down the corresponding set of differential equations and simulated this numerically on a computer. Furthermore, they included a constant influx of food molecules and an outflux of all molecule types. So, although the number of molecule *types* may be relatively small (e.g., 21 polymer types in the example above), there will be many *instances* of each type, the concentrations of which change according to the differential equations. From these simulations, they observed "a marked qualitative distinction" between networks that do and networks that do not contain an autocatalytic set (Farmer et al., 1986).

Five years later, in 1991, a pair of companion papers by Farmer and his (then) Ph.D. student Richard Bagley were published (Bagley & Farmer, 1991; Bagley et al., 1991). These papers present results from an enhanced version of the dynamic simulation model introduced earlier.

The first paper presents a detailed study of the dependency of the emergence of autocatalytic sets on network topology, (kinetic) parameter values, and the composition of the food set. As this paper concludes: "We have demonstrated that under appropriate conditions an autocatalytic set can concentrate much of the mass of its environment into a focused set, with concentrations orders of magnitude above equilibrium" (Bagley & Farmer, 1991; p. 133). However, the authors follow this up with a word of warning: "Autocatalytic metabolisms can be highly sensitive to both the topology of the reaction network and the kinetic parameters of individual reactions" (Bagley & Farmer, 1991; p. 134).

The second paper (which also includes co-author and long-time SFI associate Walter Fontana) provides preliminary results of an investigation into the possible *evolution* of autocatalytic sets. The main idea is that once an autocatalytic set has emerged and settles down into a (dynamically) stable state (i.e., no new polymer types are being produced), with some (small) probability a new polymer type is introduced into the reaction network. This simulates the occurrence of occasional "spontaneous" (uncatalyzed) reactions.

In real chemistry, a reaction can happen without a catalyst, but at a lower rate compared to when the reaction is catalyzed. However, such spontaneous reactions in the "shadow" of an existing autocatalytic set may produce a new catalyst that allows the existing set to grow and include even more polymer types (or exclude others) before it settles down into a (new) stable state. As the authors show with some preliminary numerical results: "Random variations, which play the role of mutations, are generated by spontaneous reactions. Someof these variations have no effect, and simply die out. Others have large effects, generating several new chemical species and perhaps causing others to die out, substantially altering the composition of the autocatalytic metabolism" (Bagley et al., 1991; p. 155).

Finally, it should also be noted that in 1982 the physicist Freeman Dyson independently introduced a statistical model of a population of mutually catalytic molecules undergoing random mutations (Dyson, 1982). He then studies the transition from "disorder" to "order" in his model, where his "ordered" state basically constitutes an autocatalytic set (or "catalytic network," as he calls it, after Eigen).

Dyson worked out a detailed (but highly abstract) mathematical model that gives the probability (given certain model parameters) that a population of monomers, possibly combined into polymers in an initially random way, can transform under mutation into a mutually catalytic set of molecules with high efficiency. He concludes from his model that under reasonable assumptions (such as using ten monomer species with a moderate catalytic specificity) in a population of about 2,000 monomers, it would not require "a miracle" to make the transition from the disordered to the ordered (autocatalytic) state (Dyson, 1982).

It is noteworthy that Dyson does cite and acknowledge Eigen, but not Kauffman—another indication that Kauffman's initial ideas about autocatalytic sets were probably not very widely known at that time. Dyson presents his ideas and model in more detail in a book a few years later (Dyson, 1985), but there was no further follow-up by him or anyone else.

Most of the early results described here were coherently summarized in Kauffman's first of many books (Kauffman, 1993), and in more popular form two years later (Kauffman, 1995), and clearly supported his original ideas and claims about the formation and dynamics of autocatalytic sets. They also provided a first step toward answering Eigen's criticism of the lack of evolvability in autocatalytic sets. However, another important criticism toward Kauffman's claims appeared several years later.

Criticism and resolution

A rather strong criticism of autocatalytic sets was published in 1997. In a paper on the origin of life, Israeli author Shneior Lifson includes an appendix reviewing the mathematics behind Kauffman's polymer model. As Lifson writes: "There are many problems with the model, but they need not all be discussed because of a major error which renders its conclusions wrong anyhow" (Lifson, 1997; p. 7). A few paragraphs later, Lifson points out what this major error is: "Kauffman's error was to increase M at constant P" (Lifson, 1997; p. 7).

Recall that in Kauffman's argument the maximum polymer length *M* is increased given a fixed probability of catalysis *P*, and at some point a phase transition is reached where autocatalytic sets start showing up in (random) model instances. Mathematically this might be correct, but the problem, according to Lifson, is that one should not increase *M independently* of *P*.

In the abstract model, polymers are built up of two types of monomers, 0 and 1. That means there are four possible polymers of length two: 00, 01, 10, and 11. Similarly, there are eight possible polymers of length three, 16 possible polymers of length four, and so on. In other words, each time the maximum polymer length *M* is increased by one, the total number of possible polymers doubles. And since each possible polymer can be involved in multiple reactions, the total number of possible chemical reactions more than doubles!

Such a doubling (or more) in possibilities at each step is called *exponential growth*. In other words, the number of possible reactions increases exponentially with increasing maximum polymer length *M* in Kauffman's model. As a consequence, when increasing *M* one by one, while keeping *P* fixed, the polymers in the model instances will catalyze an exponentially increasing number of reactions. And this is exactly Lifson's criticism. Mathematically it may not be a problem, but chemically it is highly unrealistic. If each polymer needs to catalyze an exponentially large number of reactions to reach the threshold where autocatalytic sets start showing up, then chemically it is not a very plausible scenario.

What should be kept constant instead of *P*, according to Lifson, is the *average* number of reactions catalyzed per molecule. In that case, it is not clear at all whether the phase transition will be reached, i.e., whether autocatalytic sets are guaranteed to emerge. Lifson therefore concludes: "Thus, the derivation of reflexively autocatalytic sets collapses" (Lifson, 1997; p. 7). Now this was an argument that was hard to ignore. But an answer was provided in an (initially) mostly unnoticed paper that appeared a few years later.

In 2000, a short and highly mathematical paper was published by Mike Steel, a mathematician from New Zealand, in which he addresses Lifson's criticism directly. First, Steel mathematically formalizes Kauffman's notion of an autocatalytic set (more on this later). Next, he considers the abstract polymer model, using the variable *n* (instead of *M*) for the maximum polymer length, and *p* (instead of *P*) for the probability of catalysis.

Steel then proves mathematically that "if each polymer catalyses on average n^2 reactions in total, then it becomes increasingly certain that the entire system of reactions is a CRA" (Steel, 2000; p. 94), where "CRA" refers to Steel's mathematical definition of an autocatalytic set. In other words, what Steel proved was that instead of an *exponential* growth rate in the (average) number of reactions catalyzed by each polymer with increasing maximum polymer length, only a *quadratic* growth rate is sufficient to get autocatalytic sets with high probability.

This was certainly a significant improvement over Kauffman's original result. However, it is still not the constant level of catalysis that Lifson insisted on. In fact, Steel also shows mathematically that if the level of catalysis, i.e., the average number of reactions catalyzed per molecule, is smaller than a particular

constant value, then the probability of autocatalytic sets existing is basically equal to zero (for increasing *n*). However, he ends his paper with the conjecture that there is some *sub*-quadratic growth rate in the level of catalysis for which there still is a high probability of autocatalytic sets existing (Steel, 2000). It took a few more years, though, before this conjecture was confirmed.

Meanwhile, additional support for Kauffman's initial results and claims came from various directions. By now several other scientists had gotten interested in autocatalytic sets as well, and started creating and studying mathematical and computational models based on Kauffman's original idea. These models all seemed to verify that autocatalytic sets do indeed arise relatively easily, and may even be evolvable.

Modeling support

In 1993, a highly mathematical paper was published by the trio Peter Stadler, Walter Fontana, and John Miller (all three long-time SFI associates). They considered a chemical system with a fixed number of molecule types, where in each possible reaction two molecule types interact to create one or more molecule types as reaction products. However, the two "reactants" are actually retained, thus playing the role of catalysts. Furthermore, a buffered food source is implicitly assumed. They then write down a set of ordinary differential equations that describes how the concentrations of the different molecule types will change over time, given a certain starting condition.

The authors show theoretically that this "catalytic network equation" has several important special cases, and also investigate numerically the effects of interconnectedness on the dynamical behavior of the network. In particular, they study a random network where each interaction (reaction) has only one unique product, and where the reaction rate constants are assigned randomly. This is similar to Kauffman's original (but more elaborate) polymer model. Indeed, the authors report that "in almost all of several hundred numerical integrations [...] the system converged to a globally stable fixed point" (Stadler et al., 1993; p. 385).

They conclude the paper by stating that their research "indicates that the typical behavior of random networks is extremely robust" (Stadler et al., 1993; p. 390) and that they "showed that the systems always reduce their dimension to a self-maintaining subset of types" (Stadler et al., 1993; p. 391). Thus, these independent investigations provided additional theoretical support for Kauffman's earlier claims about the emergence of autocatalytic sets.

A year later, in 1994, another article by Fontana (with co-author Leo Buss) was published. They introduce and study a more abstract but formal model of chemistry based on λ-calculus (lambda-calculus). In λ-calculus, objects (e.g., molecules) are defined inductively in terms of nonlinear combinations of other objects, starting from primitives. In other words, each object can also act as a function, which in the context of chemical reaction networks can be interpreted as a catalyst. The primitives (basic building blocks) define the food set.

From studying this formal model, the authors derive several main conclusions: "(i) hypercycles of self-reproducing objects arise, (ii) if self-replication is inhibited, self-maintainingorganizations arise, and (iii) self-maintaining organization, once established, can combine into higher-order self-maintaining

organizations" (Fontana & Buss, 1994a; p. 757). Furthermore, they acknowledge the relationship of their formal model and results to that of earlier work: "Our level 1 organizations recall three different lines of research. [...] The second and third research traditions are work on autocatalytic sets and on autopoietic systems" (Fontana & Buss, 1994a; p. 759). A much longer and more detailed analysis of their formal model was published that same year (Fontana & Buss, 1994b). Again, this modeling work provided additional support for Kauffman's original claims.

More supporting results were published a few years later (in 1997) by two physicists from New Zealand, Peter Wills (a former regular SFI visitor) and Leah Henderson. These authors were particularly interested in "structure-function relationships," i.e., the correspondence between polymer sequences and their catalytic properties. As already noted earlier, protein catalysts fold up in very specific three-dimensional structures, which determine their "function," i.e., which reaction(s) they are able to catalyze. Wills and Henderson (1997) incorporate such structure-function relationships into their model.

Using an abstract polymer model similar to that of Kauffman, also with ligation and cleavage reactions, these authors divide the reactions into different classes, such as the ligation of a polymer ending with 0 and one beginning with 1, or ...0 + 1... --> ...01..., represented as {0-1}. Now imagine, for example, a situation where polymers with structure 00...00 catalyze the class of reactions {1-1} (and not any other), and polymers with structure 11...11 catalyze the class of reactions {0-0} (and not any other). This would generate an autocatalytic set where all 0-polymers catalyze the formation of 1-polymers and vice versa. They then look at different such structure-function combinations to see which ones could form autocatalytic sets and which could not.

They conclude by stating that “the selection of progressively more complex collectively autocatalytic sets of polymers is possible in systems whose structure-function relationship satisfies certain constraints. By examining simple ligation/cleavage systems we illustrate what is likely to be a general precondition for the evolutionary emergence of refined biological functions: structures which carry out refined functions should be differentiated in specialized ways through the presence or absence of the refined structural features which the refined functions selectively produce." The authors realize the complicated structure of their own conclusion, as they end their paper with: "...the logic of functional evolution is strangely circular..." (Wills & Henderson, 1997).

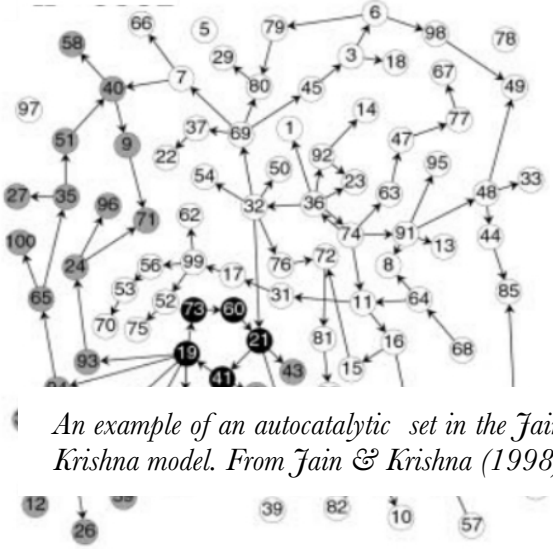
So, on the one hand they criticize Kauffman's original model for simply assigning catalysts at random, without taking structure-function relationships into account. But on the other hand, they show that even if one does take such relationships into account, autocatalytic sets still have a high probability of emerging. Their conference proceedings paper was republished more formally a few years later (Wills & Henderson, 2000), and the "Wills-Henderson model" was revisited and studied in more detail almost two decades later (Hordijk et al., 2014b).

Another theoretical study of autocatalytic sets was published in 1998 by two physicists from India, Sanjay Jain (long-time SFI associate) and Sandeep Krishna. Their model consists of a network of a fixed number of chemical species (molecule types). A link between two species means that the first species catalyzes the production of the second species from an implicitly assumed food source. Initially links are included in

the network at random, according to a given probability.

Next, with each species a dynamics is associated, where a species' concentration grows via the catalytic action of all other species that catalyze its production, and declines via a common "death rate" (or outflux). This dynamics is run until the network reaches a stable state in overall concentrations. Once such a stable state is reached, the species with the lowest concentration is replaced with a new species with completely new and random catalytic links with the other species. The network dynamics is then run again until a (new) stable state is reached, and the species with the lowest concentration is once more replaced with a new one with random catalytic links, and so on for many generations.

What Jain and Krishna observed in their model is the following. Initially the total number of links fluctuates around the expected value for a random network. But at some point, this number increases rapidly until it stabilizes again at a much higher value than it started out. They explain this rapid increase by the sudden appearance of an autocatalytic set in the network. By repeatedly removing species with low concentration (due to a lower than average connectivity) and replacing them with species with new but random links, at some point catalytic closure occurs in a subset of the species (i.e., a cycle appears in the network), giving rise to an autocatalytic set. This autocatalytic set then starts growing, as its current members will all have a high concentration and thus never be removed, until eventually the autocatalytic set encompasses (close to) the entire graph (Jain & Krishna, 1998).



An example of an autocatalytic set in the Jain-Krishna model. From Jain & Krishna (1998).

The image on the right shows an example of such an autocatalytic set existing within a network of 100 species (or nodes). The nodes in black form the "core" of the autocatalytic set, i.e., they form a cycle (or catalytic closure). The nodes in gray form the "periphery" of the autocatalytic set, i.e., those nodes of which the formation is catalyzed by other nodes in the set, but that do not feed back into the core. The white nodes are not part of the autocatalytic set. The authors then also show that it can be derived mathematically whether the network contains an autocatalytic set, and if so, which nodes are part of it. They do this by using methods from linear algebra, in particular calculating the eigenvalues and eigenvectors of the adjacency matrix that represents the network mathematically (Jain & Krishna, 1998).

Unlike the models described so far, the Jain-Krishna model explicitly includes selection. As the authors conclude: "this model provides an example of how selection for fitness at the level of individual species results, over a long time scale, in increased complexity of interaction of the collection of species as a whole. [...] when selection is operative, the system 'cashes in' upon the novelty provided by an [autocatalytic set] that arises by chance" (Jain & Krishna, 1998; p. 5687). This work was followed up with several further papers investigating the model and its results in more detail (Jain & Krishna, 2001;

2002; Giri & Jain, 2012).

In the same year (1998), a pair of papers appeared by a group of researchers from the Weizmann Institute of Science in Israel, introducing yet another model of autocatalytic sets called Graded Autocatalysis Replication Domain, or GARD (Segré et al., 1998a; 1998b). This model assumes a collection of molecule types that are chemically interconvertible via common precursors, and that are contained in a spatial "vesicle" (compartment) with a given volume. The membrane of the vesicle is permeable to the precursors, but not to other molecule types. Over time the vesicle can grow, increasing its volume to contain a larger number of molecules.

In addition, the molecule types can mutually catalyze each other's formation from the precursors. However, contrary to Kauffman's polymer model and the Jain-Krishna model, catalysis is not an "all or nothing" event, but has an efficiency associated with it, drawn from an appropriate random distribution.

This model is then analyzed mathematically in the first paper (Segré et al., 1998a) and with computer simulations in the second one (Segré et al., 1998b). In these computer simulations, a large population of small GARD vesicles is considered, each with a fixed number of molecule types randomly sampled from a larger number of chemically allowed species.Furthermore, changes in the catalytic interactions are induced by replacing one of the species by a randomly chosen one, very similar to the Jain-Krishna model. However, in the GARD model such changes are accepted only if they give rise to a vesicle with higher self-replicationefficiency.

Given the similar features of the GARD and Jain-Krishna models, similar results are obtained. Indeed, as the authors observe: "While at the initial steps in this simulated process few instances of strong mutual catalysis are present, the later stages result in the formation of a chemical network that is well-connected in terms of mutual catalysis." Furthermore: "Cycles of any size constitute powerful catalytic domains capable of catalyzing a number of other species in branched stems" (Segré et al., 1998b; p. 561). Such "catalytic domains" are similar to the cores of Jain & Krishna (1998), while the "branched stems" are similar to the periphery.

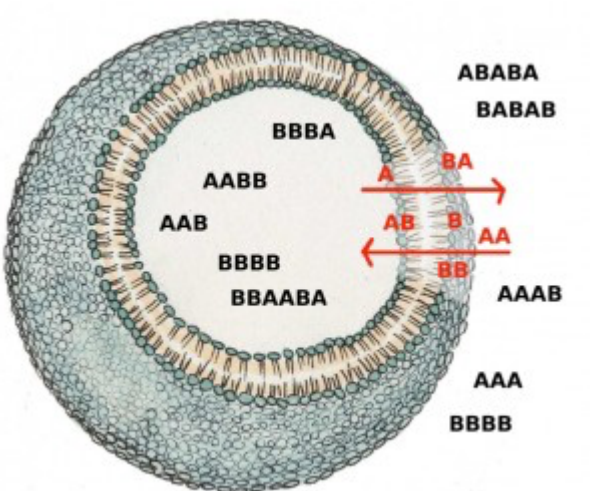
Over the years, the GARD model has been studied in great detail and in different versions, an extensive review of which can be found in Lancet et al. (2018). Together, the results from the Jain-Krishna model, the Wills-Henderson model, and the GARD model formed an important step toward answering Eigen's lack of evolvability criticism, by explicitly considering a selection process.

In 2005, Kauffman teamed up with physicists Rudolf Hanel and Stefan Thurner (long-time SFI associate) to study a mathematical model of catalytic networks very similar to that of Stadler et al. (1993) described earlier. Moreover, they also put this model and its results in the context of economics, in particular technological evolution: "We study catalytic random networks with respect to the final outcome diversity of products. [...] We demonstrate the existence of a phase transition from a practically unpopulated regime to a fully populated and diverse one" (Hanel et al., 2005; p. 1). The authors also relate their findings directly to those of Stadler et al. (1993). This work was followed up a few years later by a study

of the model in an evolutionary context (Hanel et al., 2007).

In 2011, a group of researchers in Italy led by Roberto Serra (and including Kauffman and Packard) published a paper presenting results from computer simulations similar to those of Farmer et al. (1986). These authors also used the abstract polymer model, but the dynamics were simulated using a stochastic method known as the Gillespie algorithm, rather than the deterministic differential equations method used by Farmer et al. (1986).

The authors then studied the influence of the size and composition of the food set, and of the initial molecule concentrations on the emergence of autocatalytic sets (Filisetti et al., 2011). This work was followed up by many more simulation studies, including the emergence and dynamics of autocatalytic sets in so-called protocells, i.e., small compartments with internal chemistry. In this case it was shown that synchronization takes place between the replication of the internal reaction network and that of the container, provided that the set of reactions contains an autocatalytic set (Serra et al., 2014; Villani et al., 2014; Serra & Villani, 2017).



A schematic representation of a protocell, a lipid membrane that is permeable to food molecules (monomers and dimers, in red), but not to larger polymers. From Serra et al. (2014).

However, further studies showed that this synchronization depends on several additional factors (Serra & Villani, 2019). For instance, especially when multiple autocatalytic subsets exist within a given reaction network, it depends on which molecules (from which autocatalytic subset) are coupled to the growth of the container. In some cases no synchronization occurs at all, while in other cases only one autocatalytic subset survives and synchronizes with the container growth. Furthermore, it also depends on how molecules diffuse across the membrane. If this diffusion is instantaneous (i.e., at an infinite rate), the behavior is different from when the diffusion has a finite (small) rate. In the latter case, different autocatalytic subsets may coexist within the protocell. As the authors conclude: "These observations stress the importance of a dynamic analysis whose results may lead to conclusions that are widely different from those suggested by a naive look at the static topology" (Serra & Villani, 2019; Section 5), echoing a concern already raised by Farmer et al. (1986).

Finally, in 2014, Shinpei Tanaka, Harold Fellermann, and Steen Rasmussen (former SFI associate) published a paper using a simplified version of the abstract polymer model to study the relationship between structure and selection in autocatalytic networks. Their model contains three types of reactions between polymers (or strands, as they call them): (1) decomposition of a strand into any two substrands, (2) random ligation of two strands, and (3) autocatalytic ligation. Each of these three types of reactions have their own specific rate constant, with the catalyzed rate (3) being higher than the uncatalyzed ones (1 and 2).

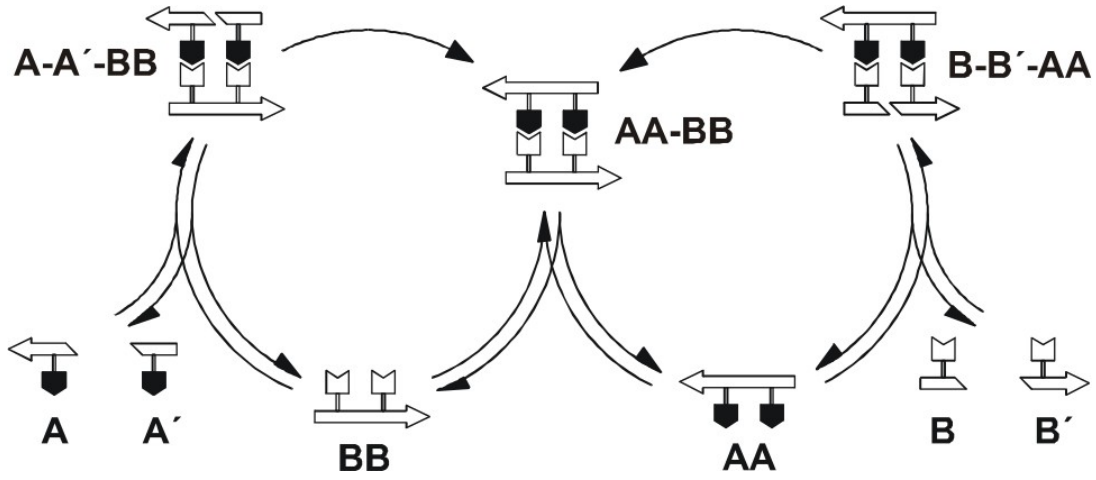
The authors then performed a theoretical analysis using a differential equation approach and a dynamical analysis using the Gillespie algorithm. What they found is that highly ordered populations with particular sequence patterns are dynamically selected out of a vast number of possible states. As the authors conclude: "to our knowledge, it has not been reported previously that the selection of specific sequence patterns arises spontaneously out of the autocatalytic dynamics. This intrinsic selection is important for the study of the origin of complex and functional polymers" (Tanaka et al., 2014; p. 28004-p6). This work was followed by a more detailed study a few years later (Fellermann et al., 2017).

In conclusion, by now a large body of work had been generated using theoretical and computational models to support Kauffman's original idea and claims on the emergence of autocatalytic sets. However, these results, although convincing, are all based on abstract models of chemical reaction networks. So what about experimental evidence?

Experimental evidence

In his book *Origins of Order*, Kauffman addresses the lack of experimental evidence for autocatalytic sets at that point: "We must consider the experimental construction of autocatalytic sets of peptides or RNA ribozymes. I suspect this construction is feasible if we are bold enough to reach the needed complexity and meet the thermodynamic requirements" (Kauffman, 1993; p. 337). As it turns out, he did not have to wait very long for such an experimental construction, and it took much less than the "needed complexity" his models had initially suggested.

A year after the publication of *Origins of Order*, a paper appeared in *Nature* reporting on the cross-catalytic replication of a pair of short nucleotide sequences (oligonucleotides) (Sievers & von Kiedrowski, 1994). This experimental system was based on two nucleotide sequences of length three (i.e., trimers), referred to as A and B. These two trimers were chosen such that they form each other's base-pair complement. Therefore, two ligated trimers of the same kind (i.e., the hexamers AA and BB) can serve as templates to which two complementary trimers can attach by forming base-pair bonds. For example, two B trimers can attach to an AA template, allowing the trimers to ligate into a fully formed BB hexamer. Note that such a ligation can also happen spontaneously, but the two trimers would have to line up exactly in the right way by chance. By attaching to a complementary hexamer, though, this ligation process is sped up.



The chemical reaction network of two cross-catalytic nucleotide-based oligomers ***AA*** *and* ***BB****. From Patzke & von Kiedrowski (2007).*

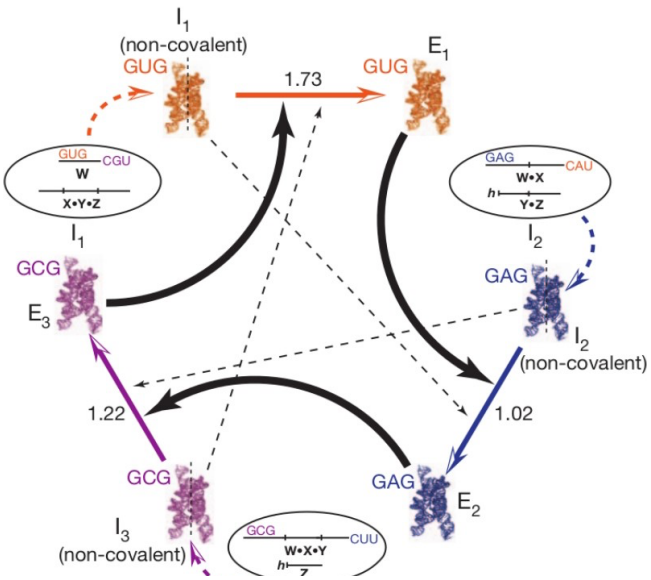
After the original template and the newly formed hexamer separate (which can be induced by subjecting the system to regular heating cycles), the original AA template is regained, plus a new BB template. In turn, this new BB template can now facilitate (speed up) the ligation of another AA template from two A trimers, and so on. In other words, the two hexamer types mutually catalyze each other's formation from smaller fragments (the food set).

This result provided the very first experimental example of a simple autocatalytic set. Indeed, the authors conclude their paper with the following statement: "Our results may have important implications for theories of the origin of life, including those that invoke self-organization of complex reaction systems involving collective replication of oligonucleotides" (Sievers & von Kiedrowski, 1994; p. 224), with a reference to Kauffman's recently published book. Kauffman had earlier promised to buy a bottle of champagne for the first person to succeed in producing such an experimental example. Living up to his promise, he and von Kiedrowski shared that bottle together.

Ten years later, a similar experimental system of two cross-catalytic nucleotide sequences (in this case of length more than 70 nucleotides each) was constructed by Kim & Joyce (2004). These RNA sequences were subsequently subjected to mutation and selection to increase their catalytic efficiency (Lincoln & Joyce, 2009).

These experimental systems of just two cross-catalytic nucleotide sequences are much smaller than the "needed complexity" that Kauffman's model had originally suggested. However, Kauffman's argument was based on random chemical reaction networks, i.e., where catalysis is assigned randomly. In contrast, the experimental examples were explicitly designed, and the nucleotide sequences were chosen in such a way that they form each other's base pair complement. Because of this, even very small systems can already become instances of an autocatalytic set.

Larger experimental systems have been constructed as well, though. In 2012, another paper appeared in *Nature*, reporting on experimental autocatalytic sets with up to 16 nucleotide sequences (Vaidya et al., 2012). These experimental systems were based on a well-studied ribozyme (i.e., a catalytic RNA sequence) of about 200 nucleotides long. By making various single-nucleotide mutations to this known ribozyme, it was possible to create systems with up to 16 different ribozymes that mutually catalyze each other's formation from smaller RNA fragments. An illustrative example of a short cycle of three mutually catalytic ribozymes is reproduced on the right.

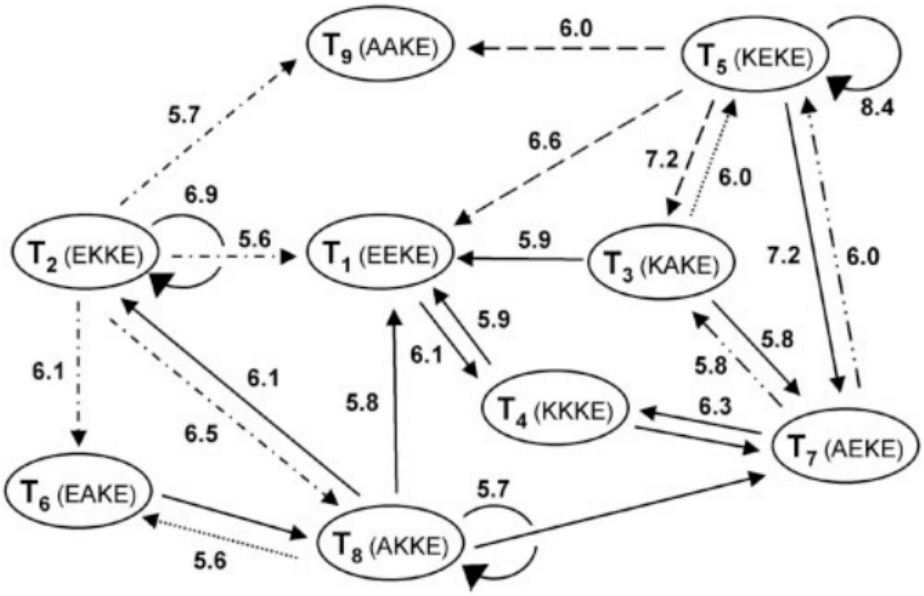


An example of an autocatalytic set of three RNA molecules. The ribozymes E_1, E_2, and E_3 are formed from smaller fragments, and each one catalyzes (curved arrows) the formation of the next one in the cycle. From Vaidya et al. (2012).

Various aspects of these autocatalytic sets were studied experimentally, including the formation of the full 16-member network from a solution containing only the initial RNA fragments (the food set). Note that in this experimental system, the catalysts are produced directly from the food set, as in the Jain-Krishna model. However, the possible construction of a more elaborate version of this RNA autocatalytic set was demonstrated later, where the production of the catalysts requires multiple reaction steps away from the food set (Arsène et al., 2018), thus more closely resembling Kauffman's abstract polymer model.

Experimental autocatalytic sets are not restricted to nucleotide sequences, though. Kauffman's original ideas were phrased mostly in terms of peptides (i.e., short proteins), and Eigen also admitted that "catalytic networks" of proteins may form quite easily. This was confirmed in 2004, when a paper was published in *PNAS* describing an experimentally constructed autocatalytic set made up of nine peptides (Ashkenasy, 2004). Similar to the RNA autocatalytic sets, each peptide in this experimental system is formed through a ligation reaction between two shorter peptide fragments, catalyzed by one or more of the other (fully formed) peptides.

Starting from a known autocatalytic peptide (32 amino acids long), several single-site mutations were introduced to generate a set of 81 sequences. The cross-catalytic efficiencies for all 81x81 pairs of sequences were then theoretically calculated. Using a given threshold value for the minimum required efficiency, this resulted in a reduced network of 25 peptides and their mutually catalytic interactions. Finally, a particular subset of nine peptides was then selected from this (theoretical) network and constructed and analyzed experimentally. A schematic overview of this 9-peptide autocatalytic set is shown in the next image.



A schematic overview of the experimental 9-peptide autocatalytic set. Arrows indicate which peptides catalyze the formation of whichother peptides. Numbers along the arrows represent the theoretically calculated catalytic efficiency. From Ashkenasy et al. (2004).

In conclusion, experimental autocatalytic sets have been constructed with either RNA or peptides. Recent results suggest they may even have formed from a mixture of the two, in a so-called nucleopeptide network (Bandela et al., 2021). However, these experimental systems all rely on shorter RNA and/or peptide fragments as their food set. Is it possible to go even further, and have a food set that consists of still simpler molecule types?

A paper published in *PNAS* in 2020 reports on an experimental autocatalytic set consisting entirely of inorganic molecules (Miras et al., 2020). These molecules are all based on molybdenum, with various auto- and cross-catalytic interactions. The results are supported by stochastic computer simulations of various dynamical aspects of the system. As the authors conclude: "The results presented here show that the formation of an autocatalytic set which embeds molecular template transfer processes can form with a simple inorganic system. [...] All previous autocatalytic sets known are derived from known biology but this study shows how autocatalytic sets, based on simple inorganic salts, can spontaneously emerge which are capable of collective self-reproduction outside of biology" (Miras et al., 2020). A remaining experimental challenge will be to bridge the gap between inorganic and polymer-based autocatalytic sets.

Formal theory

Much of the modeling and experimental work described so far was done independently by different researchers in different places, with the main connection being that they were all somehow inspired by Kauffman's original work. But recall Kauffman's definition of an autocatalytic set: "It must be the case that every member of the autocatalytic set has at least one of the possible last steps in its formation catalyzed by some member of the set, and that connected sequences of catalyzed reactions lead from the maintained 'food set' to all members of the autocatalytic set" (Kauffman, 1986; pp. 2-3). This definition

is rather informal, and perhaps even somewhat ambiguous. Indeed, different researchers have sometimes interpreted this definition in different ways.

For example, Jain and Krishna (1998) defined autocatalytic sets in terms of cores and peripheries, which can be calculated mathematically using tools from linear algebra. This method can indeed be applied when all catalysts are directly produced from the food set, as is the case in the Jain-Krishna model. However, it breaks down in the more general case where the necessary catalysts are several reaction steps away from the food set, such as in Kauffman's polymer model.

Similarly, Filisseti et al. (2011) initially looked for closed catalytic cycles ("cores") to identify autocatalytic sets. However, this does not take the "connected to a food set" part of Kauffman's definition into account. Indeed, Filisseti et al. (2011) remark that the "autocatalytic sets" they observed were not stable, and disappeared again almost as fast as they were formed. (This problem was resolved in later work, though, as is described below.)

As another example, Viadya et al. (2012) described their autocatalytic sets as instances of Eigen's hypercycles. And although similar, there are important differences between hypercycles and autocatalytic sets (more on this later), and the experimental RNA-based chemical networks were clearly instances of the latter and not the former. (This was also corrected later; see below.)

Moreover, there was no efficient computer algorithm available to identify autocatalytic sets in arbitrary reaction networks. For example, in their initial work Kauffman (1986) and Farmer et al. (1986) either identified autocatalytic sets by eye, or by the dynamical behavior resulting from their computer simulations. An autocatalytic set was assumed to be present when the system reduces its dimension to a self-maintaining subset of types, in the words of Stadler et al. (1993).

However, as mentioned earlier, to address Lifson's (1997) criticism, Steel (2000) had formalized Kauffman's original definition mathematically and unambiguously. Based on this earlier formalization, a more complete mathematical framework for autocatalytic sets was introduced (and subsequently further developed) by Wim Hordijk (former SFI graduate fellow and postdoc) and Steel (Hordijk & Steel, 2004; 2017). This framework was then also used to re-evaluate and further study many of the computational and experimental reaction networks described so far, thus providing a robust and unifying tool for detecting and studying autocatalytic sets.

First, Hordijk and Steel (2004) defined autocatalytic sets mathematically precisely as *ReflexivelyAutocatalytic and Food-generated* sets, or RAF sets. Consider a set of chemical reactions R together with the molecules (or more precisely, molecule types) involved in those reactions, and a subset of those molecule types constituting a food set F. Informally, such a reaction set R is a RAF set if:

1. Each reaction in R is catalyzed by at least one of the molecules from the set itself; and
2. Each molecule in the set can be built up from the food set F through a series of reactions from R itself.

The food set *F* is assumed to contain those molecule types that are available from the environment, and do not necessarily have to be produced by the reaction network itself. For example, in the experimental examples described earlier, the food set would consist of the smaller RNA or peptide fragments that are ligated into the longer polymers that then act as catalysts. In an origin of life setting (see below), the food set would consist of the (small) molecules that were present on the early earth.

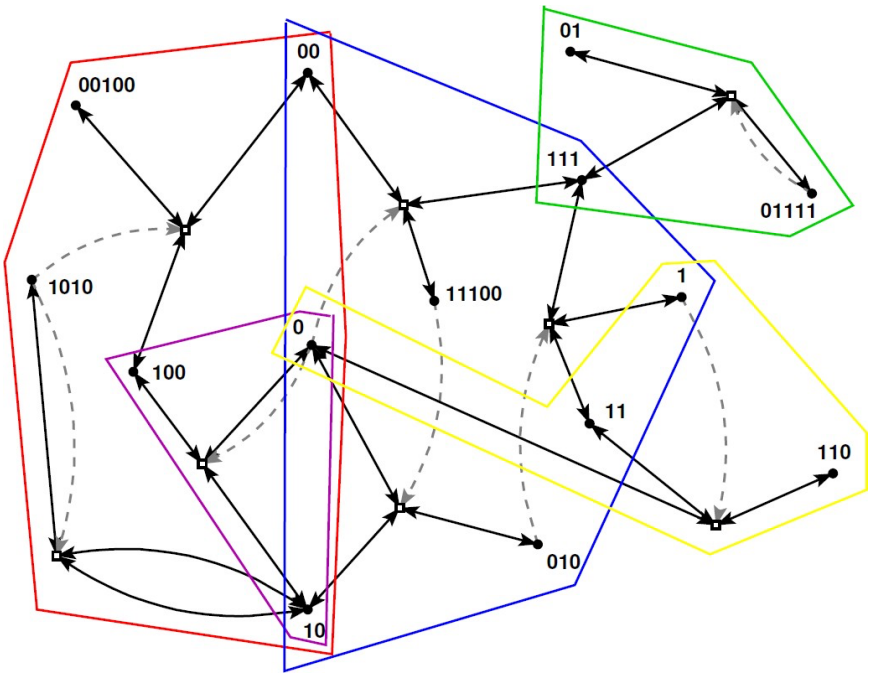
Formally, RAF sets can be defined more precisely using the mathematical notion of *closure*. Note that this is different from the notion of "catalytic closure" that Kauffman talked about. Catalytic closure refers to the occurrence of a cycle, or loop, in terms of who catalyzes the formation of whom. The mathematical notion of closure, in the context of chemical reaction networks, is defined as follows: A given set of molecules is *closed* relative to a set of reactions*R*, if after "executing" all the reactions in *R* that have all their reactants in the original set of molecules, no new molecules are produced that were not already in the original set.

With this mathematical notion the closure of a food set *F* relative to a set of reactions *R* is then defined as the set of all molecules that can be produced starting from *F* by repeated application of reactions from *R*, until no new molecules are produced. In other words, starting with a current molecule set consisting of the food set *F*, all reactions that have all their reactants in the current molecule set are allowed to be "executed," and any new molecules produced are added to the current molecule set. This is then repeated until no new molecules are produced anymore.

A RAF set can now be defined formally as a set of reactions *R* and a given food set *F*, where each reaction in *R* has *all* its reactants and *at least* one catalyst in the closure of *F* relative to *R*. This provides a mathematically precise and unambiguous definition for Kauffman's original notion of autocatalytic sets.

This formal definition also leads naturally to an efficient computer algorithm for detecting RAF sets in arbitrary reaction networks. The main idea behind this algorithm is to start with a given set of reactions *R* and a food set *F*, and then repeatedly compute the closure of *F* relative to the current reaction set *R* and remove all reactions in *R* that do not have all their reactants and at least one catalyst in the closure of *F*. At some point it will not be possible to remove any reactions from *R* anymore, and the algorithm will terminate. If upon termination the remaining set *R* is empty, then there is no RAF set in the original reaction network that was started with. If, however, the remaining set *R* is not empty, it represents the largest RAF set that is present within the original reaction network.

Note that this largest RAF set (referred to as the *maxRAF*) may contain smaller *subsets* that are RAF sets in themselves (*subRAFs*). These could either be proper subsets of each other (i.e., "nested" RAFs), or they could be completely separate or only partially overlapping RAF sets. It is possible to find such smaller RAF sets by applying the same algorithm again on the maxRAF after removing one or more reactions from it.



An example of a maxRAF that was found by the RAF algorithm in an instance of the abstract polymer model with n=5. Black dots represent molecule types, whiteboxes represent reactions. Solid arrows indicate reactants going into and products coming out of a reaction, while dashed arrows indicate catalysis. The colored outlines indicate various subRAFs within the maxRAF. The food set consists of the monomers and dimers. From Hordijk & Steel (2012).

Having this RAF algorithm available, it was possible to confirm Steel's earlier conjecture. Recall that Steel (2000) proved mathematically that a quadratic growth rate in the level of catalysis with increasing n (or maximum polymer length) in Kauffman's polymer model is sufficient to get autocatalytic sets with high probability. He then also made the conjecture that even a sub-quadratic growth rate would probably suffice.

After running the RAF algorithm for several weeks on a large computer cluster on many (random) instances of the polymer model for various values for the parameters n and p, a clear picture emerged. Even a *linear* growth rate in the level of catalysis is sufficient to get RAF sets with high probability (Hordijk & Steel, 2004). In fact, extrapolating from these computer simulation results, for n up to about 50 (i.e., a maximum polymer length of up to 50), the polymers do not need to catalyze more than two reactions each (on average) to get these RAF sets. Chemically this is highly plausible, and thus refutes Lifson's (1997) original criticism. Inspired by these computational results, Steel and co-author Elchanan Mossel subsequently managed to also prove theoretically that a linear growth rate in the level of catalysis indeed suffices to get autocatalytic sets with high probability (Mossel & Steel, 2005).

Note that in the RAF algorithm, in computing the closure of the food set catalysts are not considered. In other words, a reaction is allowed to be "executed" when all its reactants are present in the current molecule set. Chemically this makes sense, as a chemical reaction can always proceed as long as all the reactants are present, although it may happen at a much lower rate compared to when the reaction is

catalyzed. In a RAF set, it is required that eventually the set produces all its own catalysts, and all reactions are catalyzed. But in some cases this may actually require an uncatalyzed (spontaneous) reaction to happen at first.

As an example, consider the experimental system of Sievers and von Kiedrowski (1994), with two nucleotide hexamers catalyzing each other's ligation from trimers (assuming the trimers constitute the food set). When starting with the food set (trimers) only, there are initially no catalysts available. So, one of the ligation reactions will have to happen spontaneously, i.e., without being catalyzed. Once this has happened though, the first catalyst is available to catalyze the other ligation reaction, at which point the full autocatalytic (RAF) set comes into existence (in terms of dynamics). So, there may be an initial waiting time for the required spontaneous reaction(s) to happen, depending on the rate at which they happen uncatalyzed, but once they have occurred, all reactions within the RAF set can proceed catalyzed.

This points to an important issue regarding network topology and dynamics. Even if a RAF set is present in the underlying reaction network, it is not guaranteed that it will be realized dynamically, depending on how many spontaneous reactions are required initially and at what rate these might occur. Also, differences in the efficiencies of different catalysts may play an important role dynamically. As Farmer et al. (1986; p. 58) already stated, "a consideration of kinetics is essential for a realistic assessment of the potential to generate autocatalytic sets," a point re-iterated by the group of Italian researchers (Serra & Villani, 2017).

Mossel and Steel (2005) made an explicit distinction between RAFs that might require spontaneous reactions to be dynamically realized, and RAFs that do not require such spontaneous reactions. The latter they referred to as CAFs, for *Constructively Autocatalytic and Food-generated* sets. CAFs can be dynamically constructed from the food set immediately, with all catalysts present when they are needed, without requiring any spontaneous reactions. CAFs are thus a special subset of RAFs (i.e., a CAF is also a RAF, but the reverse is not generally true). Note that for a RAF to also be a CAF, at least one of the food molecules will need to be a catalyst.

In hindsight it turns out that in their early work, Kaufmann (1986) and Farmer et al. (1986) only considered CAFs. Interestingly, for CAFs the original criticism by Lifson (1997) actually *does* apply. As Mossel and Steel (2005) proved theoretically, for the more general RAF sets a linear growth rate in the level of catalysis suffices to get these RAF sets with high probability (in random instances of the polymer model), but for the more strict CAF sets an *exponential* growth rate in the level of catalysis is required! In other words, not allowing spontaneous reactions (at least initially) makes the occurrence of autocatalytic sets much less likely.

There are, however, at least two reasons why RAF sets seem a more natural choice to consider than CAF sets. First, spontaneous reactions *do* happen in real chemistry. In fact, catalysts do not "invent" new chemistry, they simply speed up reactions that already occur naturally anyway. And second, referring back to Eigen's original criticism, spontaneous reactions play an important role in the potential evolvability of autocatalytic sets. This evolvability issue is addressed in more detail below.

With the formal RAF framework (including the efficient RAF algorithm) available, many of the already mentioned computational and experimental systems have been re-evaluated and studied in more detail. For example, on the computational side the problem of the instability of the catalytic cycles in the initial work of Filisseti et al. (2011) was resolved by incorporating the RAF algorithm to look for true autocatalytic sets, i.e., also taking into account the food-generated part (Filisetti et al., 2014). Furthermore, the main results of Jain and Krishna (1998) were re-evaluated using the formal RAF framework (Hordijk, 2016; Steel et al., 2019).

On the experimental side, the RNA autocatalytic sets of Vaidya et al. (2012) were studied in more detail using the formal RAF framework (Hordijk & Steel, 2013; Hordijk et al., 2014a). This not only led to the correction of some minor errors in the presentation of the original results, but also to further insights into the emergence and structure of these RNA autocatalytic sets that would have been difficult to obtain through experiments alone. Similarly, the experimental peptide autocatalytic set of Ashkenasy et al. (2004) was studied in more detail with the formal RAF framework, also leading to additional insights and possible avenues for further experiments (Hordijk et al., 2018b).

Perhaps even more importantly, though, the RAF framework has also been applied to study real biological reaction networks.

Metabolic networks

Recall that life's complexity, i.e., it being more than the sum of its constituent molecules, was illustrated with a (hypothetical) experiment involving *E. coli* bacteria. *E. coli* is a species of bacteria that is part of your normal gut flora. It produces vitamin K, and can help fight off other, harmful bacteria. *E. coli* is one of the most studied micro-organisms, and its reconstructed metabolic network is the most complete of any living species. An organism's metabolic network comprises the set of all chemical reactions and interactions that occur within that organism to turn food into biomass and usable energy.

It seems a natural question to ask whether the metabolic network of living organisms contain autocatalytic (RAF) sets: "However, despite their appeal, the relevance of RAFs for real biochemical networks that exist in nature has, so far, remained virtually unexplored" (Sousa et al., 2015; p. 1). When molecular biologists Bill Martin and Filipa Sousa teamed up with Hordijk and Steel, they entered this virtually unexplored terrain by searching for autocatalytic sets (using the RAF algorithm) in the metabolic network of *E. coli*.

Even though the reconstructed metabolic network of *E. coli* is the most complete of any organism, it still has missing data. But a bigger hurdle had to be overcome. In an organism's metabolism, nearly all reactions are catalyzed by enzymes, i.e., long proteins. These protein catalysts do not form randomly though, they are explicitly encoded by genes. In other words, even though the basic building blocks for enzymes (i.e., individual amino acids) are produced by the metabolic network itself, generating the specific protein sequences for the network's catalysts requires the genetic system (including DNA, RNA, and the ribosome, or the "translation device"). Therefore, a metabolic network with protein catalysts would not contain a RAF, as these catalysts require something more than the metabolic network itself to

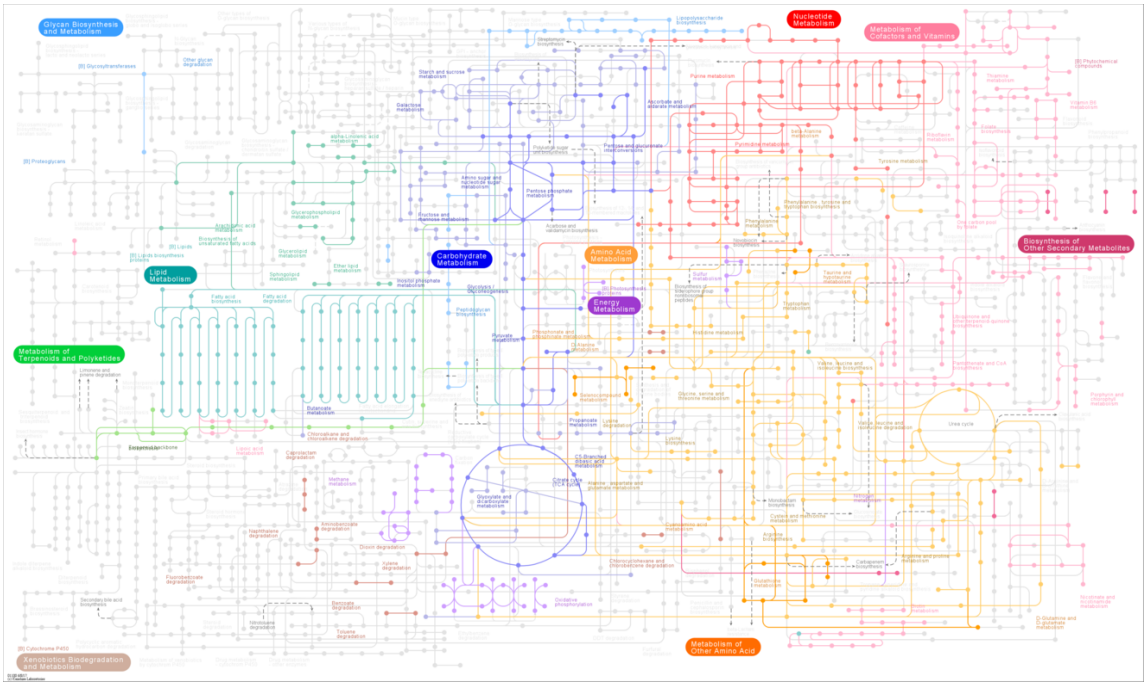
be produced.

The solution Sousa et al. (2015) came up with was to consider *cofactors* as catalysts. Most enzymes, next to consisting of a long protein, contain one or more small molecules (often referred to as cofactors), such as various metals like iron, zinc, or magnesium, or organically produced molecules like ATP, flavin, or CoA. In fact, it is often the cofactor that really performs the catalysis. The complicated three-dimensional structure of the protein largely serves to hold everything (the reactants and the cofactor catalyst) in the right place. This way, the protein makes the cofactor a more efficient catalyst. And in many cases, an enzyme's protein structure has evolved to catalyze only one or a few particular metabolic reactions (but doing so very efficiently), whereas the cofactor alone can generally catalyze many different reactions (although less efficiently).

Indeed, the RAFs found in the metabolic network of *E. coli* only have a small number of (cofactor) catalysts, just over 40, that together catalyze the close to 1,800 reactions in the network. As the authors observe: "The critical role of cofactors in the *E. coli* RAFs might point to an interesting aspect of early chemical evolution. We see here that the size, hence in some respects the complexity, of RAFs within the *E. coli* metabolic network are dependent upon cofactors: a small number of catalysts that promote a large number of reactions each" (Sousa et al., 2015; p. 16).

Moreover, these RAFs contain a modularity that corresponds closely to functional groups in metabolism in general. This modularity was discovered by investigating the influence of single molecules or reactions on the size of the RAF. For example, after removing a single molecule or reaction, the RAF algorithm can be applied again to see how much the remaining RAF is reduced compared to the original RAF. As it turns out, there are many molecules/reactions that have a small impact on the size of the RAF, and a few molecules/reactions that have a large impact (Sousa et al., 2015).

As the authors conclude: "The existence of RAF sets within a microbial metabolic network indicates that RAFs capture properties germane to biological organization at the level of single cells" (Sousa et al., 2015; p. 1). This is a rather crucial indication that was a first of its kind, and one that supports Kauffman's original claim that autocatalytic sets are an essential underlying property of all living systems.



The metabolic network of E. coli, which contains a large autocatalytic set when cofactors are considered as the catalysts. Image produced with iPath (Darzi et al., 2018).

The main reason these authors chose *E. coli* for their study is that, as mentioned, it has the most complete reconstructed metabolic network of any organism. So it was an obvious place to start. However, *E. coli* is a highly evolved bacterium, with a large metabolic network. How about "simpler" microbes, especially those that are believed to be closer to the very first living organisms that appeared on Earth around four billion years ago?

Joana Xavier, also working with Bill Martin, decided to search for autocatalytic sets in the metabolic networks of another bacterium, *Moorella thermoacetica*, and of one archaeon, *Methanococcus maripaludis*. These microbes represent primitive lineages that live on the simplest source of carbon and energy known, and are assumed to be closely related to some of the earliest living organisms. Not only did Xavier et al. (2020) show that the metabolic networks of these organisms do indeed contain RAF sets, but also that their *intersection* contains one. This autocatalytic set is interpreted as the RAF of LUCA, the *Last Universal Common Ancestor*: "RAFs uncover elements of metabolic evolution that precede the divergence of archaea and bacteria from the LUCA" (Xavier et al., 2020).

As with the original *E. coli* study of Sousa et al. (2015), cofactors (rather than complete enzymes) were used as the catalysts in this more recent study. Surprisingly, though, even with these small-molecule catalysts (several of which are naturally occurring inorganic elements), the "RAF of LUCA" is actually able to produce some amino acids and nucleotides, the basic building blocks of proteins and DNA/RNA (Xavier et al., 2020). Therefore, these results could have important implications for our understanding of the actual *origin* of life, to which we will return shortly, after having a more detailed look at the possible *evolvability* of autocatalytic sets.

Evolvability

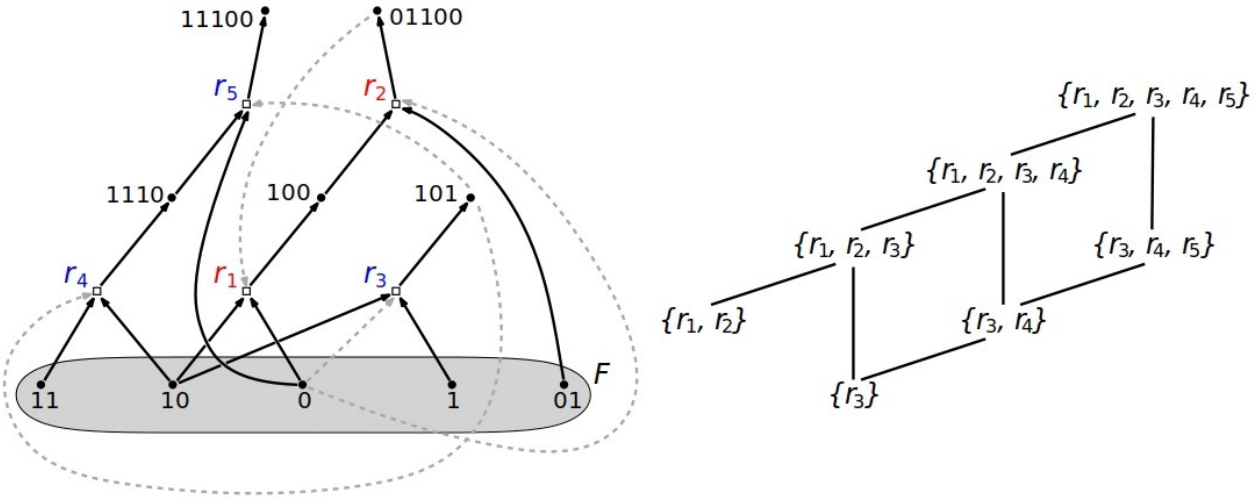
Recall Eigen's (1971) criticism that autocatalytic sets would not be evolvable. This criticism was partially resolved in some of the early theoretical and computational studies on autocatalytic sets. For example, Bagley et al. (1991) allowed occasional spontaneous reactions to happen (simulated by introducing new polymers into the system at random), which sometimes substantially altered the composition of already established autocatalytic sets. Furthermore, results from the Wills-Henderson model (Wills & Henderson, 1997), the Jain-Krishna model (Jain & Krishna, 1998), and the GARD model (Lancet et al., 2018), as described above, also show various ways in which autocatalytic sets can potentially evolve. But the most convincing evidence for the possible evolvability of autocatalytic sets was provided by a group of researchers led by evolutionary biologist Eőrs Szathmáry (and also including Kauffman).

First, these researchers make a distinction between the "core" of an autocatalytic set (i.e., a closed catalytic loop), and its "periphery" (i.e., catalyzed reactions branching out from the core), just as Jain & Krishna (1998) had defined earlier. Next, as Bagley et al. (1991) had also done, they allow spontaneous (uncatalyzed) reactions to happen with low probability. This occasionally generates a new catalyst that could give rise to an entirely new core coming into existence. Finally, they assume that the autocatalytic sets are contained within compartments (such as lipid membranes) that grow and divide, distributing the internal molecules between the offspring compartments randomly. Thus, in their own words, "'Mutation' happens either when uncatalyzed reactions result in the emergence of a novel core, or when molecular components of a viable core are stochastically lost after compartment splitting" (Vasas et al., 2012; p. 10).

Their computer simulations (using a version of the abstract polymer model) then lead them to state: "We conclude that only when a chemical reaction network consists of many such viable cores, can it be evolvable. When many cores are enclosed in a compartment there is competition between cores within the same compartment, and when there are many compartments, there is between-compartment competition due to the phenotypic effects of cores and their periphery at the compartment level. Acquisition of cores by rare chemical events, and loss of cores at division, allows macromutation, limited heredity, and selectability, thus explaining how a poor man's natural selection could have operated prior to genetic templates" (Vasas et al., 2012; p. 1).

However, this still left open the question of how many autocatalytic cores one could expect to exist within a given reaction network. After all, evolution thrives on diversity, so the more cores that are potentially available, the more "evolvable" the system could be. Independently, but at the same time, Kauffman had also teamed up with the duo Hordijk and Steel, publishing a paper that same year that provided at least a partial answer to this question.

First recall that the RAF algorithm finds the maxRAF, i.e., the largest RAF that is present in a given reaction network. However, as mentioned, a maxRAF may contain smaller subsets that in themselves are also RAF sets. Indeed, using an example of a small (five-reaction) RAF set that was found by their RAF algorithm in an instance of the polymer model, Hordijk et al. (2012) show that this RAF set consists of several smaller RAF subsets (subRAFs). Moreover, these subRAFs form a hierarchical structure known as a *partially ordered set* (or *poset*) in mathematical terms. The example maxRAF and its poset of subRAFs are reproduced below.



Left: A maxRAF as found in an instance of the abstract polymer model. Right: The poset of subRAFs within the maxRAF. Modified from Steel et al. (2019).

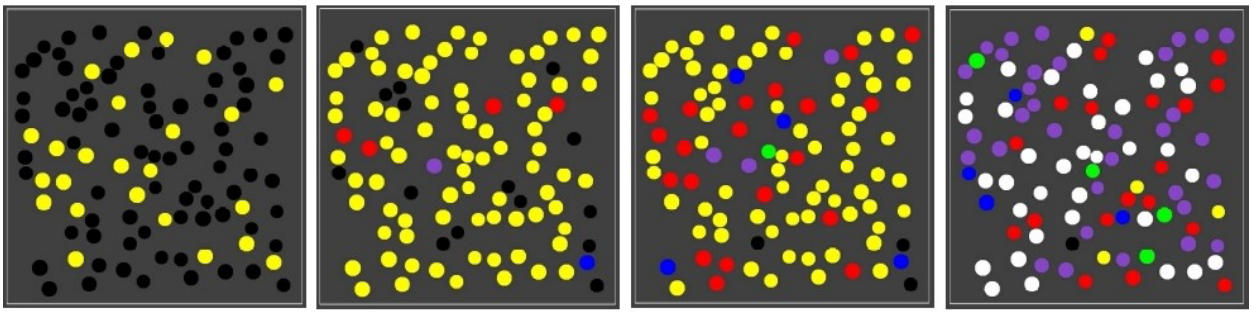
Next, note that the two subRAFs at the left-bottom of the poset ($\{r_1,r_2\}$ and $\{r_3\}$) do not contain any smaller RAF subsets. They are therefore called *irreducible* RAFs, or irrRAFs. It can be shown (by construction) that a given maxRAF can, in principle, contain an exponentially large number of irrRAFs. In particular, Hordijk et al. (2012) provide an example of a hypothetical RAF set consisting of $2k$ reactions that contains 2^k irrRAFs.

Finally, the notion of an irrRAF corresponds closely to that of a (viable) core of Vasas et al. (2012). In other words, a given reaction network that contains a large enough (max)RAF could thus (potentially) contain a very large number of autocatalytic cores, i.e., sufficient diversity to enable a rudimentary evolutionary process to take place. In later work, empirical estimates of the actual number of irrRAFs in instances of the abstract polymer were presented (Steel et al., 2013; Hordijk et al., 2015).

The potential for evolvability of autocatalytic sets was further illustrated in collaboration with a group of researchers in the UK (including Fellermann), who had recently developed an agent-based simulation toolkit called *Simbiotics* for studying the collective behavior of single-celled organisms such as bacteria (Naylor et al., 2017). Realizing that this toolkit could also be used to study the emergence and dynamics of autocatalytic sets in populations of compartments, together these scientists developed a simulation module to do just that (Hordijk et al., 2018a).

The overall setup is as follows. A population of compartments exists in a two-dimensional spatial environment that has a constant influx of food molecules (monomers and dimers) that diffuse throughout the system, and a constant outflux of all molecule types. However, compartments are permeable to food molecules but not to larger molecule types, just as in the earlier simulations of Villani et al. (2014). So, when autocatalytic sets are formed inside compartments they can be maintained, whereas in the outside environment they would dilute away. The authors then used an instance of the polymer model that contains several smaller autocatalytic subsets, and watched what happened over time.

As expected, different combinations of autocatalytic subsets (subRAFs) started to appear in different compartments. The emergence of these particular subRAFs requires one or more spontaneous reactions to happen (at low rates), which are stochastic events. So, in one compartment one particular subRAF may appear at some point (color it blue), while in anothercompartment another subRAF may appear at some later time (color it red). Sometime later still, a compartment already containing a "blue" subRAF may also acquire a "red" subRAF, turning the compartment "purple" (i.e., both red and blue together), and so on. The image below presents four snapshots over time, clearly showing how the compartment colors(representing the various combinations of subRAFs they contain) change over time.



Four snapshots over time (from left to right) from a dynamical simulation with a population of compartments. From Hordijk et al. (2018a).

As the authors conclude: "Our simulations show that the main requirements for autocatalytic sets to be evolvable are met when encapsulating them into compartment populations: the existence of different combinations of autocatalytic subsets [...] in a population of compartments, giving rise to different 'cell types' and competition between them" (Hordijk et al., 2018a; p. 12). They then also study several additional scenarios, such as one subRAF generating a molecule that is "toxic" to another subRAF, or certain "inducer" molecules being allowed to diffuse between compartments, thereby increasing the chances that other compartments will also produce certain autocatalytic subsets.

Very recently, a group of researchers in Paris examined the evolvability of autocatalytic sets experimentally (Ameta et al., 2021). Using variants of the RNA autocatalytic sets originally reported by Vaidya et al. (2012), the dynamical behavior of these sets was studied using microfluidics. This technology consists of microscopic water droplets ("microdroplets") suspended in oil, as a way of experimentally simulating compartments.

As the authors conclude: "Indeed, in autocatalytic networks, Darwinian evolution is in principle possible despite the absence of template-based replication, but relies on the appearance (due to rare reactions or environmental changes) of catalytic species that are sustained by autocatalysis and modulate the composition of a pre-existing network" (Ameta et al., 2021). However, they explicitly add that true evolvability depends on different trade-offs within these networks and their dynamics, such as trade-offs between growth and variation, and between variation and robustness. But overall these recent results form an exciting experimental validation of the earlier computational studies on the evolvability of autocatalytic sets.

So, after Lifson's criticism about the required level of catalysis had already been resolved, Eigen's criticism of the lack of evolvability is now also resolved. Not only are autocatalytic sets likely to emerge (even for very moderate and realistic levels of catalysis), they are also (potentially) evolvable. As already alluded to above, all this could have important implications for our understanding of the *origin* of life.

Origin of life

In his initial work, Kauffman already concluded: "...the formation of autocatalytic sets of polypeptide catalysts is an expected emergent collective property of sufficiently complex sets of polypeptides, amino acids, and other small molecules. This could have substantial implications for the origin of life" (Kauffman, 1986; p. 11-12). This conclusion was repeated by Farmer et al.: "Our results suggest that autocatalytic properties may have played a major role in supplying the complex chemical prerequisites needed for the origin of life" (Farmer et al., 1986; p. 62).

For the past (more than) three decades, the dominant paradigm in origin of life research has been the RNA world hypothesis. The main idea behind this hypothesis is that one of the first stages in the origin of life consisted solely of self-replicating RNA molecules. Based on the (then) recent discovery that RNA molecules can catalyze chemical reactions such as splicing of other RNA molecules, it was suggested that protein enzymes would not have been necessary for the origin of life. In the words of Walter Gilbert, the Nobel laureate who popularized the idea: “One can contemplate an RNA world, containing only RNA molecules that serve to catalyse the synthesis of themselves” (Gilbert, 1986).

After more than 30 years of efforts, though, nobody has found or constructed a single self-replicating RNA molecule, i.e., an RNA molecule that catalyzes its own template-directed formation, nucleotide by nucleotide. Partial success has been achieved (Wochner et al., 2011), but a fully self-replicating RNA molecule remains elusive.

Furthermore, given its elusiveness, a true template-based self-replicating RNA molecule would likely require a very specific sequence of nucleotides. This means that mutant variants of such a highly specific self-replicator will almost certainly lose this self-replicating ability very quickly again. This, in turn, means that they would not be very evolvable (or not at all).

In contrast, as has been described here at length, collections of molecules that *mutually* catalyze each other's formation from smaller building blocks have been shown to emerge quite easily in models of chemical reaction networks. Furthermore, they have been constructed experimentally, and they have been shown to be evolvable through a form of "compositional inheritance."

Bagley and Farmer actually ended their paper with an explicit stab at the (then) recently introduced notion of an RNA world: "This model adds support to the idea that the emergence of a metabolism may have preceded the emergence of a self-replicator based on templating machinery" (Bagley & Farmer, 1991; p. 134). The "compositional inheritance" of autocatalytic sets (as opposed to genetic inheritance) may only allow for a limited form of evolution, but could very well have been an important step toward true open-ended evolution. As Vasas et al. state: "However, a viable core constitutes one bit of heritable information and therefore the number of possible selectable attractors is relatively small, meaning that

autocatalytic networks may not be able to sustain open-ended evolution. While we think this to be the case, the potential role of these autocatalytic networks as a route to nucleotide-based template self-replicating systems should not be underestimated" (Vasas et al., 2012; p. 12).

Moreover, the work on autocatalytic sets in real metabolism (Sousa et al., 2015; Xavier et al., 2020) shows that the emergence of autocatalytic sets may not need protein enzymes, but could happen with basic cofactors as the initial catalysts (including naturally occurring inorganic ones). Given that the autocatalytic sets in even the most primitive lineages known are already able to produce some nucleotides and amino acids (Xavier et al., 2020), one could easily imagine small RNAs and proteins forming, incorporating the initial catalysts as their cofactors, thus making them more specific and efficient catalysts. This opens the way for even more complex polymers to be formed, creating even more specific and efficient catalysts, and so on, in an upward spiral of complexity and diversity.

So, an alternative picture to the overly simplistic (and seemingly unrealistic) RNA world hypothesis is starting to emerge, one that is based on the spontaneous formation and subsequent further evolution (and complexification) of autocatalytic sets. And these autocatalytic sets are not necessarily restricted to just one type of molecule (such as just RNA, or just proteins), but most likely involved a mix of small molecules (cofactors and other inorganics) together with small nucleotide, amino acid, and fatty acid polymers.

Autocatalytic sets, in their original conception, were largely inspired by the origin of life problem. How does chemistry become biology? How do chemical reaction networks acquire self-sustaining and self-reproducing abilities, and how do they evolve to become more complex and diverse? Autocatalytic sets are indeed assumed to have played a fundamental role in these processes. However, more recently the concept of autocatalytic sets has been extended to networks beyond chemistry and the origin of life. In particular, it has been argued that it is also relevant to ecology, economics, and perhaps even cognition.

Beyond chemistry

There are two main arguments behind the claim that autocatalytic sets may be relevant to areas beyond chemistry such as ecosystems and economies. The first argument is that the formal framework of autocatalytic sets, RAF theory, is purely based on abstract networks. The second argument is that the chemical notion of catalysis can be generalized to other settings.

Recall that the definition of a RAF set was explicitly stated in chemical terms such as molecule types, chemical reactions, and catalysis. However, as the various examples have shown, such RAF sets can generally be represented by a network, i.e., nodes representing either molecule types or chemical reactions, and arrows (links) between nodes representing either reaction inputs and outputs or catalysis. Such a network representation is more abstract, and many of the RAF theorems are purely based on the mathematical properties of such networks. In other words, many of the theorems and results from RAF theory are independent of the fact that these networks represent chemical reaction networks.

From a mathematical point of view, it does not matter what the nodes and arrows in the network

represent, as long as we know something about, for example, what the probability distribution of the catalysis arrows is in random instances of these networks. The RAF definition and resulting theorems could just as well have been stated purely in terms of abstract nodes and arrows. As a consequence, any system that can be represented by a similar type of network could also be considered in terms of RAF sets.

In an economy, goods are being transformed into other goods, which is generally referred to as a "production function." For example, wood and nails as inputs can be transformed into a table as output. Thus, such an economic production function can be viewed as the equivalent of a chemical reaction. Furthermore, there is the equivalent of a "food set" consisting of raw materials, such as iron ore (which is used to produce nails) and trees (which are used to produce pieces of wood).

Similarly, ecosystems are often represented by so-called "food-webs," i.e., a network of "who eats whom." In other words, some species serve as inputs (i.e., the ones that are being eaten) to "produce" other species (i.e., the ones that are doing the eating). This can also be considered the equivalent of a chemical reaction. The species at the lowest level of this network (the ones that can live on whatever the environment offers) constitute the food set.

But what about catalysis? Is it possible to find the equivalent of catalysts in an economy or ecosystem? If one is willing to consider the notion of catalysis in a more abstract way as well, then the answer is a clear "yes." Recall that catalysts do not "invent" new chemistry, but mostly increase the rates at which possible chemical reactions happen, without being used up in those reactions. In the same way, some goods in an economy increase the rates of certain production functions, or some species in an ecosystem increase the rate of production of certain other species, but without being used up.

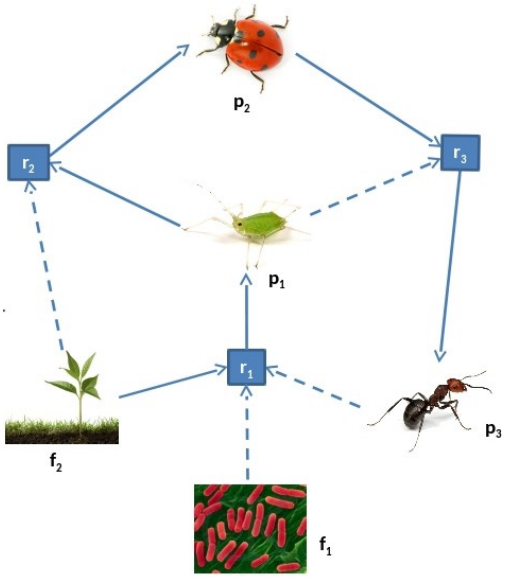
For example, by using a hammer the rate of producing tables from wood and nails can be significantly increased. However, the hammer (unlike the wood and nails) is not used up in this process. The same hammer can be used again to make the next table, and the next one, and so on. Many other goods, such as conveyor belts, printing presses, and computers, can thus also be considered catalysts.

Similarly, certain tree species provide safe nesting space for certain bird species, thus allowing those bird species to (re)produce faster than they would otherwise. However, a tree is not used up (eaten) in this process. The same tree can provide nesting space to a new generation of birds the next breeding season. Many other species, such as bees pollinating flowers, little fish cleaning the teeth of larger fish, or gut bacteria that help us digest our food, can thus also be considered catalysts.

Furthermore, all these economic and ecological catalysts are themselves products within their respective networks. So the question arises quite naturally whether an economy or an ecosystem could perhaps also be considered as a self-sustaining autocatalytic set in which the elements (goods or species) mutually catalyze each other's production. This question has recently started to be investigated more seriously by

several researchers in the contexts of economics and ecology (Hanel et al., 2005; Cazzolla-Gatti et al., 2017; 2018) and even in cognition (Gabora & Steel, 2017; 2020).

An example of a (proposed) autocatalytic set in ecology is reproduced on the right. The food set consists of bacteria (f_1) and plants (f_2). Aphids (p_1) eat plant sap (r_1), which lacks certain essential amino acids. However, the aphids have acquired gut bacteria that produce those missing amino acids. Thus, these bacteria act as a catalyst for the "production" of aphids (they are not consumed by the aphids). Aphids, however, are eaten (r_2) by ladybugs (p_2). Plants infested with aphids producea certain chemical that attracts ladybugs, and thus act as a catalyst for the "production" of ladybugs. In response, aphids produce a sweet substance that is harvested by ants (p_3), which in return provide protection against ladybugs, which they sometimes attack and eat (r_3). Aphids thus act asa catalyst for the "production" of ants, closing the catalytic loop.



Example of an ecosystem autocatalytic set. From Cazzolla Gatti et al. (2018).

Back to Table of Contents

IV. Related Ideas

Of course autocatalytic sets are not one of a kind. In fact, as was already pointed out earlier, other (similar) notions have been proposed independently, such as Eigen's (1971) hypercycles and Dyson's (1982; 1985) mutually catalytic sets of proteins.

A *hypercycle* is a collection of self-replicating macromolecules (i.e., they each catalyze their own formation from a food source), where each molecule additionally also catalyzes the replication of the next molecule, in a closed cyclic manner (Eigen, 1971). The idea behind this is that if, for example, due to mutations, one macromolecule loses its ability to self-replicate, it can still be formed through a reaction catalyzed by the preceding molecule in the cycle, thereby maintaining the integrity of the system as a whole.

Recall that Eigen introduced this concept after rejecting the idea of autocatalytic sets due to their perceived lack of evolvability. However, a hypercycle is actually a very specific *instance* of an autocatalytic set. In particular, it is an autocatalytic set where each molecule catalyzes exactly two reactions: its own formation and that of the next molecule in the cycle. Chemically this seems a rather difficult requirement though. Indeed, there are currently no known chemical examples of hypercycles, either natural or experimentally constructed. Unfortunately, though, there is still much confusion between strict hypercycles and the more general notion of autocatalytic sets. This "conceptual error" was addressed in detail by Szathmáry (2013) and also briefly by Hordijk (2017).

A formalism known as *metabolism-repair systems*, or (*M*,*R*) systems, closely related to that of autocatalytic sets but more abstract, was introduced by Robert Rosen already back in the 1950s and 60s (Rosen, 1991). However, it never gained much attention, as it has been difficult to understand due to its rather abstract formulation. In 2010, a group of researchers, mostly from Chile, explicitly pointed out the close connection between (*M*,*R*) systems and RAF sets (Jaramillo et al., 2010). In particular, they try to make the concept of (*M*,*R*) systems more clear by rephrasing it in terms of RAF sets: "An important unresolved matter is to make explicit how Rosen's equations can be fulfilled using concepts and definitions imported from RAF sets" (Jaramillo et al., 2010; p. 99). However, they also note some differences. For example, in (*M*,*R*) systems all catalysts are required to be strictly produced by reactions from the system itself, whereas in RAF sets catalysts could, in principle, also come from the food set, not necessarily being produced by any of the reactions from the set. As a consequence, (*M*,*R*) systems (like hypercycles) are specific *instances* of RAF sets, but in some cases additional features need to be taken into account to conform to Rosen's formalism.

Several years earlier, Pier Luigi Luisi had noted, in passing, a similarity between autocatalytic sets and the notion of autopoietic systems (Luisi, 2003). The theory of *autopoietic systems* (Varela et al., 1974; Luisi, 2003) tries to explain life as a functionally closed and self-sustaining chemical system. In other words, autopoietic systems organize the production of their own components in such a way that these components are continuously regenerated and therefore maintain the chemical network processes that produce them. The notion of a *boundary* (such as a cell membrane) is essential in this model, physically separating the system from its environment, but allowing certain nutrients to enter and waste products to leave. However, this boundary layer must be produced by the system itself, and in turn promote the further production of its constituent components (Luisi, 2003).

Although boundaries are not considered explicitly in RAF theory, they can be dealt with quite easily. In particular, a boundary can be considered an additional catalyst, given that it keeps all internal molecules in close enough proximity so that they can chemically interact, rather than dilute away. This idea, and the similarity between autopoietic systems and RAF sets, was explored further by Hordijk & Steel (2015).

Luisi also discusses similarities between autopoietic systems and *chemotons* (Luisi, 2003). The *chemoton model*, by the Hungarian chemical engineer Tibor Gánti, has two complementary chemical reaction networks within a self-generated boundary ("membrane system"): a metabolic network ("cyclic subsystem") and an informational network ("genetic subsystem"), where the metabolic subsystem is (at least partly) controlled by the genetic subsystem, and the metabolic subsystem in turn provides the basic building blocks for both the genetic subsystem and the membrane (Gánti, 1975; 2003). The system as a whole is thus self-sustaining (given a food source), in an autocatalytic way, providing a formal model of cellular life as we know it.

In Smith et al. (2014) a partitioned polymer model was studied in the context of RAF sets, where reactions can only involve molecule types from one of two partitions (e.g., either only RNA or only peptides), but catalysis can be both within and across partitions. This study showed that the existence of RAF sets is equally likely (and for similar levels of catalysis) as in a standard non-partitioned polymer model. Thus,

autocatalytic systems with an explicit distinction between a metabolic and a genetic network, as in the chemoton model, can also be dealt with in terms of RAF sets.

Furthermore, a precise mathematical correspondence between RAF sets and chemical organizations was derived recently (Hordijk et al., 2018c). *Chemical organization theory* (COT) is an alternative formal framework for defining and studying closed and self-maintaining reaction networks (Dittrich & Speroni di Fenizio, 2007). However, chemical organizations do not require an explicit notion of catalysis or a food set. Nevertheless, there isa direct correspondence between the two theories, and the mathematical representation of one theory can be readily converted into one or the other. In particular, there is a direct correspondence between chemical organizations and so-called "closed" RAF sets, which are exactly the relevant RAF subsets in the context of evolvability of autocatalytic sets (Hordijk et al., 2018a).

Finally, autocatalytic sets, with their explicit catalytic closure, have been considered a specific (and concrete) instance of the more general phenomenon of functional organization through *constraint closure* (Montévil & Mossio, 2015; Moreno, 2016; Lehman & Kauffman, 2021).

In conclusion, there are several models that, in various but related ways, try to capture the self-dependent and self-maintaining complex nature of life. However, many of these models remain mostly abstract, or do not have any real chemical examples. In contrast, RAF theory successfully combines three important results:

1. A solid mathematical foundation (including dynamic simulations and an efficient detection algorithm),
2. Various experimentally constructed chemical examples, and
3. Formally verified biological examples.

Furthermore, there are close correspondences between RAF theory and the other formalisms, where one can often be expressed in terms of the other. As such, RAF theory can serve as a unifying general framework for all of these related models and ideas.

Back to Table of Contents

V. The Future

This research synthesis has provided a general introduction and overview of the notion of autocatalytic sets, and the scientific progress that has been made over the past 50 years.This work has generated an impressive body of both theoretical and experimental results, coming from a wide variety of disciplines, from pure mathematics and computer science to chemistry and biology, and even economics and cognition.

However, much remains to be done. For instance, experimentally constructed examples of autocatalytic sets exist, consisting either of inorganic molecules or of biological polymers such as RNA or peptides. What still needs to be done is to bridge the gap between these two classes of examples, i.e., to show how organic (RNA or peptide) sets can evolve from inorganic ones, or how mixed (organic and inorganic) sets

can emerge.

Furthermore, the potential evolution of autocatalytic sets needs to be shown more explicitly experimentally. As mentioned earlier, efforts are underway using microfluidics technology, which seems to be a promising direction. This can be aided by further computer simulations on autocatalytic sets in protocells.

The eventual goal will be to have autocatalytic sets spontaneously emerging and evolving in controlled laboratory settings, starting from a relatively simple food set such as a combination of inorganic and small organic (e.g., cofactor) molecules. This could have important applications in, for example, medicine and synthetic biology. From a scientific point of view, it may provide a significant step toward understanding how life as we know it originated, or how *artificial* life could be generated and evolved from scratch.

Also, applying the notion of autocatalytic sets beyond chemistry to economics and ecology still needs a more formal foundation. Furthermore, some examples based on real-world data sets would be very helpful. Such a formalism may help in identifying weak spots in these types of systems. For example, it would allow for identification of the elements (e.g., firms or species) that have the largest impact on the size of the autocatalytic set upon their removal from the set.

Over the years, more and more scientists (and non-scientists) have become interested in autocatalytic sets, as this review has highlighted. In fact, it has become almost impossible to keep track of everything that is being done and published in this field. And even Kauffman himself, now in his 80s, is still actively involved in some of the ongoing work. With these combined efforts, across the many different disciplines involved, autocatalytic sets have become an increasingly important and relevant example of complexity at the interface of chemistry and biology, and potentially also beyond. Starting from an initial idea 50 years ago, it has become a research field on its own, with an exciting and promising future. Let's hope the next 50 years will bring that promise to full fruition.

Back to Table of Contents.

Acknowledgements

I would like to thank the many colleagues with whom I have been privileged to work together on this highly interesting topic. Special thanks go to Stuart Kauffman and Mike Steel for their support, long-standing collaboration, and friendship over the years. I also thank the Santa Fe Institute (in particular Jennifer Dunne) for the opportunity to contribute this paper, the writing of which was funded through a grant (ID# 61988) from the John Templeton Foundation.

VI. References

Ameta, S., Arsène, S., Foulon, S., Saudemont, B., Clifton, B. E., Griffiths, A. D. & Nghe, P. Darwinian properties and their trade-offs in autocatalytic RNA networks. *Nature Communications* 12:842, 2021.

Arsène, S., Ameta, S., Lehman, N., Griffiths, A. D. & Nghe, P. Coupled catabolism and anabolism in autocatalytic RNA sets. *Nucleic Acids Research* 46(18):9660-9666, 2018.

Ashkenasy, G., Jegasia, R., Yadav, M. & Ghadiri, M. R. Design of a directed molecular network. *PNAS* 101(30):10872-10877, 2004.

Bagley, R. J. & Farmer, J. D. Spontaneous emergence of a metabolism. In: Langton C. G., Taylor, C., Farmer, J. D. & Rasmussen, S. (eds.), *Artificial Life II*. Addison-Wesley, pp. 93-140, 1991.

Bagley, R. J., Farmer, J. D. & Fontana, W. Evolution of a metabolism. In: Langton C. G., Taylor, C., Farmer, J. D.& Rasmussen, S. (eds.), *Artificial Life II*. Addison-Wesley, pp. 141-158, 1991.

Bandela, A. K., Wagner, N., Sadihov, H., Morales-Reina, S., Chotera-Ouda, A., Basu, K., Cohen-Luria, R., Escosura, A. de la & Ashkenasy, G. Primitive selection of the fittest emerging through functional synergy in nucleopeptide networks. *PNAS*, 2021 (in press).

Cazzolla Gatti, R., Hordijk, W. & Kauffman, S. A. Biodiversity is autocatalytic. *Ecological Modelling* 346:70-76,2017.

Cazzolla Gatti, R., Fath, B., Hordijk, W., Kauffman, S. A. & Ulanowicz, R. Niche emergence as an autocatalytic process in the evolution of ecosystems. *Journal of Theoretical Biology* 454:110-117, 2018.

Dittrich, P. & Speroni di Fenizio, P. Chemical organization theory. *Bulletin of Mathematical Biology* 69(4):1199-1231, 2007.

Dyson, F. J. A model for the origin of life. *Journal of Molecular Evolution* 18:344-350, 1982.

Dyson, F. J. *Origins of Life*. Cambridge University Press, 1985.

Eigen, M. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58(10):465-523, 1971.

Erdős, P. & Rényi, A. On random graphs. *Publicationes Mathematicae* 6:290-297, 1959.

Erdős, P. & Rényi, A. On the evolution of random graphs. *Publications of the Mathematical Institute of theHungarian Academy of Sciences* 5:17-61, 1960.

Farmer, J. D., Kauffman, S. A. & Packard, N. H. Autocatalytic replication of polymers. *Physica D* 22:50-67, 1986.

Fellermann, H., Tanaka, S. & Rasmussen, S. Sequence selection by dynamical symmetry breaking in an autocatalytic binary polymer model. *Physical Review E* 96:062407, 2017.

Filisetti, A., Graudenzi, A., Serra, R., Villani, M., De Lucrezia, D., Fuchslin, R. M., Kauffman, S. A., Packard, N. &Poli, I. A stochastic model of the emergence of autocatalytic cycles. *Journal of Systems Chemistry* 2:2, 2011.

Filisetti, A., Villani, M., Damiani, C., Graudenzi, A., Roli, A., Hordijk, W. & Serra, R. On RAF sets and autocatalytic cycles in random reaction networks. *Communications in Computer and Information Science* 445:113-126, 2014.

Fontana, W. & Buss, L. W. What would be conserved if the tape were played twice? *PNAS* 91:757-761, 1994a.

Fontana, W. & Buss, L. W. The arrival of the fittest: Toward a theory of biological organization. *Bulletin of Mathematical Biology* 56:1-64, 1994b.

Gabora, L. & Steel, M. Autocatalytic networks in cognition and the origin of culture. *Journal of Theoretical Biology* 431:87-95, 2017.

Gabora, L. & Steel, M. A model of the transition to behavioural and cognitive modernity using reflexively autocatalytic networks. *Journal of Royal Society Interface* 17:20200545, 2020.

Gánti, T. Organization of chemical reactions into dividing and metabolizing units: The chemotons. *BioSystems* 7:15-21, 1975.

Gánti, T. *The Principles of Life*. Oxford University Press, 2003.

Gilbert, W. The RNA world. *Nature* 319:618, 1986.

Giri, V. & Jain, S. The origin of large molecules in primordial autocatalytic reaction networks. *PLoS ONE* 7(1):e29546, 2012.

Hanel, R., Kauffman, S. A. & Thurner, S. Phase transition in random catalytic networks. *Physical Review E* 72:036117, 2005.

Hanel, R., Kauffman, S. A. & Thurner, S. Towards a physics of evolution: Critical diversity dynamics at the edges of collapse and bursts of diversification. *Physical Review E* 76:036110, 2007.

Hordijk, W. Evolution of autocatalytic sets in computational models of chemical reaction networks. *Origins of Life and Evolution of Biospheres* 46:233-245, 2016.

Hordijk, W. Autocatalytic confusion clarified. *Journal of Theoretical Biology* 435:22-28, 2017.

Hordijk, W., Naylor, J., Krasnogor, N. & Fellermann, H. Population dynamics of autocatalytic sets in a compartmentalized spatial world. *Life* 8:33, 2018a.

Hordijk, W., Shichor, S. & Ashkenasy, G. The influence of modularity, seeding, and product inhibition on peptide autocatalytic network dynamics. *ChemPhysChem* 19:2437-2444, 2018b.

Hordijk, W., Smith, J. I. & Steel, M. Algorithms for detecting and analysing autocatalytic sets. *Algorithms for Molecular Biology* 10:15, 2015.

Hordijk, W. & Steel, M. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of Theoretical Biology* 227(4):451-461, 2004.

Hordijk, W. & Steel, M. Autocatalytic sets extended: Dynamics, inhibition, and a generalization. *Journal of Systems Chemistry* 3:5, 2012.

Hordijk, W. & Steel, M. A formal model of autocatalytic sets emerging in an RNA replicator system. *Journal of Systems Chemistry* 4:3, 2013.

Hordijk, W. & Steel, M. Autocatalytic sets and boundaries. *Journal of Systems Chemistry* 6:1, 2015.

Hordijk, W. & Steel, M. Chasing the tail: The emergence of autocatalytic networks. *BioSystems* 152:1-10, 2017.

Hordijk, W., Steel, M. & Dittrich, P. Autocatalytic sets and chemical organizations: Modeling self-sustaining reaction networks at the origin of life. *New Journal of Physics* 20:015011, 2018c.

Hordijk, W., Steel, M. & Kauffman, S. A. The structure of autocatalytic sets: Evolvability, enablement, and emergence. *Acta Biotheoretica* 60(4):379-392, 2012.

Hordijk, W., Vaidya, N. & Lehman, N. Serial transfer can aid the evolution of autocatalytic sets. *Journal of Systems Chemistry* 5:4, 2014a.

Hordijk, W., Wills, P. R. & Steel, M. Autocatalytic sets and biological specificity. *Bulletin of Mathematical Biology* 76(1):201-224, 2014b.

Jain, S. & Krishna, S. Autocatalytic sets and the growth of complexity in an evolutionary model. *Physical Review Letters* 81(25):5684-5687, 1998.

Jain, S. & Krishna, S. A model for the emergence of cooperation, interdependence, and structure in evolving networks. *PNAS* 98(2):543-547, 2001.

Jain, S. & Krishna, S. Large extinctions in an evolutionary model: The role of innovation and keystone species. *PNAS* 99(4):2055-2060, 2002.

Jaramillo, S., Honorato-Zimmer, R., Pereira, U., Contreras, D., Reynaert, B., Hernández, V., Soto-Andrade J., Cárdenas, M. L., Cornish-Bowden, A. & Letelier, J. C. (M,R) systems and RAF sets: Common ideas, tools and projections. In: *Proceedings of the ALife XII Conference*, pp. 94-100, 2010.

Kauffman, S. A. Cellular homeostasis, epigenesis and replication in randomly aggregated macromolecular systems. *Journal of Cybernetics* 1(1):71-96, 1971.

Kauffman, S. A. Autocatalytic sets of proteins. *Journal of Theoretical Biology* 119:1-24, 1986.

Kauffman, S. A. *The Origins of Order*. Oxford University Press, 1993.

Kauffman, S. A. *At Home in the Universe*. Oxford University Press, 1995.

Kim, D.-E. & Joyce, G. F. Cross-catalytic replication of an RNA ligase ribozyme. *Chemistry & Biology* 11:1505-1512, 2004.

Lancet, D., Zidovetzki, R. & Markovitch, O. Systems protobiology: Origin of life in lipid catalytic networks. *Journalof the Royal Society Interface* 15:20180159, 2018.

Lehman, N. E. & Kauffman, S. A. Constraint closure drove major transitions in the origins of life. *Entropy* 23:105, 2021.

Lifson, S. On the crucial stages in the origin of animate matter. *Journal of Molecular Evolution* 44:1-8, 1997.

Lincoln, T. A. & Joyce, G. F. Self-sustained replication of an RNA enzyme. *Science* 323:1229-1232, 2009.

Luisi, P. L. Autopoiesis: A review and a reappraisal. *Naturwissenschaften* 90:49-59, 2003.

Miras, H. N., Mathis, C., Xuan, W., Long, D.-L., Pow, R. & Cronin, L. Spontaneous formation of autocatalytic sets with self-replicating inorganic metal oxide clusters. *PNAS* 117(20):10699-10705, 2020.

Mitchell, M. *Complexity—A Guided Tour*. Oxford University Press, 2009.

Montévil, M. & Mossio, M. Biological organisation as closure of constraints. *Journal of Theoretical Biology* 372:179-191, 2015.

Moreno, A. Some conceptual issues in the transition from chemistry to biology. *History and Philosophy of theLife Science*s 38:16, 2016.

Mossel, E. & Steel, M. Random biochemical networks: The probability of self-sustaining autocatalysis. *Journal ofTheoretical Biology* 233(3):327-336, 2005.

Naylor, J., Fellermann, H., Ding, Y., Mohammed, W. K., Jakubovics, N. S., Mukherjee, J., Biggs, C. A., Wright, P. C. & Krasnogor, N. Simbiotics: A multiscale integrative platform for 3D modeling of bacterial populations. *ACS Synthetic Biology* 6(7):1194-1210, 2017.

Patzke, V. & von Kiedrowski, G. Self-replicating systems. *Arkivoc* 2007(5):293-310, 2007.

Rosen, R. *Life Itself*. Columbia University Press, 1991.

Segré, D., Lancet, D., Kedem, O. & Pilpel, Y. Graded autocatalysis replication domain (GARD): Kinetic analysisof self-replication in mutually catalytic sets. *Origins of Life and Evolution of Biospheres* 28:501-514, 1998.

Segré, D., Pilpel, Y. & Lancet, D. Mutual catalysis in sets of prebiotic organic molecules: Evolution through computer simulated chemical kinetics. *Physica A* 249:558-564, 1998.

Serra, R., Filisetti, A., Villani, M., Graudenzi, A., Damiani, C. & Panini, T. A stochastic model of catalytic reaction networks in protocells. *Natural Computing* 13(3):367-377, 2014.

Serra, R. & Villani, M. *Modelling Protocells*. Springer, 2017.

Serra, R. & Villani, M. Sustainable growth and synchronization in protocell models. *Life* 9(3):68, 2019.

Sievers, D. & von Kiedrowski, G. Self-replication of complementary nucleotide-based oligomers. *Nature* 369:221-224, 1994.

Skoultchi, A. I. & Morowitz, H. J. Information storage and survival of biological systems at temperatures near absolute zero. *Yale Journal of Biology and Medicine* 37:158-163, 1964.

Smith, J. I., Steel, M. & Hordijk, W. Autocatalytic sets in a partitioned biochemical network. *Journal of Systems Chemistry* 5:2, 2014.

Sousa, F. L., Hordijk, W., Steel, M. & Martin, W. F. Autocatalytic sets in *E. coli* metabolism. *Journal of Systems Chemistry* 6:4, 2015.

Stadler, P. F., Fontana, W. & Miller, J. H. Random catalytic reaction networks. *Physica D* 63:378-392, 1993.

Steel, M. The emergence of a self-catalysing structure in abstract origin-of-life models. *Applied Mathematics*

Letters 3:91-95, 2000.

Steel, M., Hordijk, W. & Smith, J. I. Minimal autocatalytic networks. *Journal of Theoretical Biology* 332:96-107, 2013.

Steel, M., Hordijk, W. & Xavier, J. C. Autocatalytic networks in biology: Structural theory and algorithms. *Journal of the Royal Society Interface* 16:20180808, 2019.

Szathmáry, E. On the propagation of a conceptual error concerning hypercycles and cooperation. *Journal of Systems Chemistry* 4:1, 2013.

Tanaka, S., Fellermann, H. & Rasmussen, S. Structure and selection in an autocatalytic binary polymer model. *EPL* 107:28004, 2014.

Thurner, S., Hanel, R. & Klimek, P. *Introduction to the Theory of Complex Systems*. Oxford University Press, 2018.

Vaidya, N., Manapat, M. L., Chen, I. A., Xulvi-Brunet, R., Hayden, E. J. & Lehman, N. Spontaneous network formation among cooperative RNA replicators. *Nature* 491:72-77, 2012.

Varela, F., Maturana, H. & Uribe, R. Autopoiesis: The organization of living systems, its characterization and a model. *BioSystems* 5:187-195, 1974.

Vasas, V., Fernando, C., Santos, M., Kauffman, S. A. & Sathmáry, E. Evolution before genes. *Biology Direct* 7:1, 2012.

Villani, M., Filisetti, A., Graudenzi, A., Damiani, C., Carletti, T. & Serra, R. Growth and division in a dynamic protocell model. *Life* 4:837-864, 2014.

Wills, P. & Henderson, L. Self-organisation and information-carrying capacity of collectively autocatalytic sets of polymers: Ligation systems. In: Bar-Yam, Y. (ed.), *Proceedings of the International Conference on Complex Systems*. New England Complex Systems Institute, 1997.

Wills, P. & Henderson, L. Self-organisation and information-carrying capacity of collectively autocatalytic sets of polymers: Ligation systems. In Bar-Yam, Y. (ed.), *Unifying Themes in Complex Systems*, Volume I. Westview Press, pp. 613-623, 2000.

Wochner, A., Attwater, J., Coulson, A. & Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science* 332:209-212, 2011.

Wolfenden, R. & Snider, M. J. The depth of chemical time and the power of enzymes as catalysts. *Accounts of Chemical Research* 34(12):938-945, 2001.

Xavier, J. C., Hordijk, W., Kauffman, S. A., Steel, M. & Martin, W. F. Autocatalytic chemical networks at the origin of metabolism. *Proceedings of the Royal Society B* 287:20192377, 2020.