A DIAGRAMMATIC AND CATEGORICAL APPROACH TO THE OUROBOROS EQUATION

JUAN-CARLOS LETELIER, JORGE SOTO-ANDRADE, AND AMARANTA VALDÉS-ZORRILLA

1. INTRODUCTION

Biological systems are extremely difficult to analyze and common mathematical tools (like systems of ordinary differential equations) are quite inadequate to fathom the central core of biological phenomena. Interestingly, a pure algebraic approach, based on Category Theory, seems possible if we shift our attention from the metaphor that living systems are optimizing machines (the current paradigm in biology) to the one of self constructing machines (a new paradigm that slowly arose almost 60 years ago with the work of (Rosen, 1958).

In this paper, reporting on work in progress, we aim at putting forth a tentative categorical approach to metabolic systems, in terms of (proto)metabolic graphs and the associated metabolic (pathway) categories, as mathematical objects that represent metabolism (and other self - constructing systems), because they possess the property of *interconvertibility*. In effect molecules in any cell, are produced by complex biochemical reaction networks, where these molecules are not only substrates and products but also operators as they rule (channel) the dynamics of the network. Thus in a metabolic graph or a metabolic category, objects can become arrows, and arrows turn out to be objects.

We claim moreover that solutions to the mythical "Ouroboros equation"

$$f(f) = f$$

the ultimate self-referential equation (Soto-Andrade et al., 2011), could be fathomed as metabolic graphs and categories, besides being constructed by limiting procedures which embody the infinite regress that Rosen tried to avoid in his formalisation of living systems as (M,R) systems (loc. cit.).

Recall that the Ouroboros, the ancient symbol of the snake eating its own tail, is indeed a metaphor of choice to represent self-reference and circularity (Maturana and Varela, 1973, 1980; Soto-Andrade et al., 2011). In Ouroboros equation, f (supposedly a function) applies to itself, as an argument, the result being again f. So f plays simultaneously the roles of argument, function and value! The Ouroboros lurks indeed in metabolic systems, which are inherently circular: in any cell the processes that construct the cell are implemented by components made by the cell itself. Metaphorically, we can thus state that metabolism acting upon metabolism produces metabolism. So we could write a suggestive equation that requires little explanation in order to make sense:

$$metabolism(metabolism) = metabolism$$

This insight is important as it provides a direction in which to think the mathematical nature of f. It appears indeed that f should be, at least, a complex entity similar to the operation of a complex chemical reaction network but also to a system defining the identity of its constituents (in this case the identity is produced by the particular set of chemical reactions where each component participates). Thus it does not require a great leap of faith (or understanding) to assume that f should be a generalization of the basic notion of a function. Now, since Eilenberg and Mac Lane, we know that a category (seen as a bunch of arrows, each one endowed with a source and a target, which are eventually concatenable)

does provide such a generalization. The idea to take advantage of category theory in biology can be traced back to the earlier work of Rosen (Rosen, 1958, 1959).

Remarkably, motivation to consider Ouroboros equation did not arise from everyday mathematics proper. It arose from various fields ranging from Logic and Computer Science to Theoretical Biology. We have in fact called *"Ouroboros avatars"* (Soto-Andrade et al., 2011), the various ways in which Ouroboros equation has emerged in different domains. So we have avatars of Ouroboros in Logic (Löfgren, 1968; Scott, 1972, 1973), Hyperset Theory (Aczel, 1988), Cognitive Sciences (Kampis, 1995; Kauffman, 1987), Computer Science and Informatics (Scott, 1972; Kampis, 1995; Milner, 2006), Systems Theory and Theoretical Biology (Rosen, 1991; Soto-Andrade and Varela, 1984; Maturana and Varela, 1980; Letelier et al., 2005, 2006), among others, as surveyed in Soto-Andrade et al. (2011). A surprising fact is the similarity of methods of constructing solutions to the Ouroboros equation, developed in fields apparently as unrelated as logic (Scott, 1972, 1973) and metabolic systems theory (Letelier et al., 2005, 2006), motivated by the construction of actual mathematical models for untyped lambda calculus and virtual infinite regress in metabolic systems, respectively.

To begin with, it can be proved that Ouroboros is not an oxymoron (Soto-Andrade et al., 2011), i.e. that the existence of an object f such that $f(f) = f$, belonging to its own domain and range, is not logically inconsistent (Löfgren, 1968; Kampis, 1995). Instead, it turns out that an atomically self-reproducing entity can be axiomatized, and in this sense it really does exist (Löfgren, 1968). Our viewpoint is however that Ouroboros lives indeed outdoors, with respect of our usual logical - mathematical realm, but just outside, by the doorstep, so to say, so that it could be *approximated* stepwise " from within". This intuition that we explain below, has been captured to a great extent, in different guises, in Scott (1972, 1973); Soto-Andrade and Varela (1984), in Varela's further work (Varela and Goguen, 1978) and in Letelier et al. (2006).

2. Ouroboros as the limit of an infinite regress in *(M,R)* systems

We recall now Rosen's synthetic insights regarding metabolic circularity, that he developed completely independently of Scott (for a comprehensive survey of references about Rosen's work see Letelier et al. (2006))

2.1. **Rosen's *(M,R)* systems and infinite regress.**

In his formalism of *(M,R)* systems, the collective action of the thousands of catalysts in a metabolic network *M* coalesces into a single mapping f from A, the collection of all sets of reactants, to B, the collection of all sets of products, which transforms inputs $a \in A$ into outputs $b = f(a) \in B$. So a may be interpreted as the set of all LHS (left hand sides) of biochemical reactions in *M* and b as the set of all RHS (right-hand sides) of those reactions. In our interpretation what Rosen does is describing category by a "global map" $f : A \longrightarrow B$ obtained by putting together all its arrows, all its sources and all its targets; a rather misleading simple description indeed.

Now, since in any metabolic system, catalysts need to be regenerated or replaced by the system, Rosen introduced a replacement procedure Φ, which from a suitable $b = f(a) \in B$ as input, reproduces f according to $\Phi(b) = f$. Because Φ is actually selecting from the relatively large set $H(A, B) \subset Map(A, B)$, of all possible metabolisms (supposedly the set of morphisms from A to B in a suitable concrete category), a specific f such that $f(a) = b$, using $b \in B$ as an input, Rosen calls it a *selector*. Thus, according to Rosen, the replacement procedure Φ is embodied as a morphism from B to $H(A, B)$. In our interpretation, Φ would appear simply as a mapping from the set of all arrows of the category $\mathcal{C}_M$ to its set of targets. Rosen works in fact in a concrete category of structured sets,

which is implicitly assumed to be a closed category, like the category of finite dimensional vector spaces over a fixed base field.

According to Rosen however, an *(M,R)* system has the following algebraic description based on two morphisms $f \in H(A,B)$, $\Phi \in H(B,H(A,B))$ acting in synergy:

$$A \xrightarrow{f} B \xrightarrow{\Phi} H(A,B)$$
$$a \longmapsto f(a) = b \longmapsto \Phi(b) = f$$

But then the system should be able to replace the replacer Φ of f, by means of a procedure (a morphism again) $\beta : H(A,B) \longrightarrow H(B,H(A,B))$ such that $\beta(f) = \Phi$, etc. ...This is called a *replicative (M,R)* system in Rosen's terminology. This property is also referred as *organizational invariance* in Cárdenas et al. (2010).

The big question arises then, how can this be, without implying an infinite regress?

Rosen's solution to avoid infinite regress, is to "shunt" it, positing that the equation $\Phi(b) = f$ is to have only one solution Φ (a most demanding constraint indeed) so that the mapping β sends f to this unique selector Φ. In other words, β is just the inverse of the "evaluation at b" operator ev_b (acting on functions whose domain contains b) so that no further procedure is needed to construct β itself. It is in this sense that Rosen claims that his construction solves the problem of infinite regress although we was unable to give concrete examples where his hypothesis was fulfilled.

In Rosen's notations the operation of an *organizationally invariant (M,R)* system can therefore be viewed as three morphisms (f,Φ,β) acting in synergy:

$$A \xrightarrow{f} B \xrightarrow{\Phi} H(A,B) \xrightarrow{\beta} H(B,H(A,B))$$
$$f(a) = b, \qquad \Phi(b) = f, \qquad \beta(f) = \Phi.$$

where β is the inverse of the "evaluation at b" operator ev_b. Here his $H(X,Y)$ notation suggests that the set of possible metabolisms with reactants A and products B should be conceived as the morphisms of a suitable (metabolic) category.

Our remark now, is that if instead of shunning infinite regress, as Rosen did, we look at it "face to face", a recursive construction emerges, which in categorical terms may be described as follows:

If you have a morphism $f : A \to B$, in a category $\mathcal{C}$ (of structured sets and structure preserving mappings, for the time being), can you find a new morphism $f_1 : B \to C$ such that for a suitable $a \in A$ you have $f_1(f(a)) = f$?

In the category of sets the answer is obviously yes, but in more general categories, with less morphisms, this is not so clear. When trying to answer this question for a given closed category $\mathcal{C}$, where morphism sets $H(A,B)$ are objects of $\mathcal{C}$, and evaluation maps are morphisms, we would take C to be $H(A,B)$, and we would just need to find a morphism f_1 from B to $C = H(A,B)$, such that $f_1(b) = f$.

Recall that Rosen, to avoid infinite regress, posited the uniqueness of such a function f_1, called Φ in his setup (Rosen, 1991; Letelier et al., 2006). In the category of sets this is clearly impossible, unless B is a singleton. Nevertheless, if we have a category of structured sets whose sets of morphisms $H(X,Y)$ are much smaller than $Map(X,Y)$, existence of f_1 may become less obvious and uniqueness may become more possible. We may hope then for the existence of a *turning point* in the choice of our category $\mathcal{C}$, at which the sets of morphisms $H(X,Y)$ would have the right size so as to have simultaneously existence and uniqueness of our morphism f_1.

However, if we look at infinite regress face to face and we do not care about uniqueness, we could continue our above construction forever, as in Soto-Andrade and Varela (1984) under a mild hypothesis of existence of our replacing morphisms f_1, in the framework of a concrete' category of structured sets $\mathcal{C}$, to wit:

Definition 1. *Let $\mathcal{C}$ be a closed category of structured sets, so that all "evaluation at x" mappings $ev_x : f \mapsto f(x)$ $(x \in A)$ from $H(A,B)$ to B are morphisms of $\mathcal{C}$, i. e. $ev_x \in H(H(A,B),B)$.*

We say that $\mathcal{C}$ is a replacement category *(also called a Rosenean category) if the following holds.*

Given any morphism $f \in H(A,B)$, we can choose $a \in A$ such that

- there exists a morphism $f_1 \in H(B, H(A,B))$, such that $f_1(f(a)) = f$ (we say then that a is an $f-generic$ element),

- there exists a morphism $f_2 : H(H(A,B), H(B, H(A,B)))$, such that $f_2(f_1(f(a))) = f_1$ (i.e. f(a) is $f_1-generic$), and so on...

By way of example, it is easy to check that the following categories are replacement categories:

- The category of finite dimensional vector spaces and linear maps.
- The category of finite commutative groups and group homomorphisms.

2.2. **Infinite regress leads to Ouroboros.**

Proposition 2. *Assume that the category $\mathcal{C}$ is a replacemente category. Then we can construct the following infinite sequence of morphisms (and objects) issued from any morphism $C_0 \overset{\Phi_0}{\to} C_1$ in $\mathcal{C}$:*

$$C_0 \overset{\Phi_0}{\to} C_1 \overset{\Phi_1}{\to} C_2 \overset{\Phi_2}{\to} C_3 \overset{\Phi_3}{\to} \dots \overset{\Phi_{n-1}}{\to} C_n \overset{\Phi_n}{\to} C_{n+1} \overset{\Phi_{n+1}}{\to} \dots$$

$$c_0 \to c_1 \to c_2 \to c_3 \to \dots \to c_n \to c_{n+1} \to \dots$$

satisfying the following:

$$C_2 = H(C_0, C_1), \dots, C_{n+1} = H(C_{n-1}, C_n)$$

and $\Phi_n \in H(C_n, C_{n+1}) = C_{n+2}$,

$\Phi_1(\Phi_0(c_0)) = \Phi_0$ *for a suitable* $c_0 \in C_0$,

$\Phi_n(c_n) = c_{n+1} \in C_{n+1}$ $(n \geq 0)$ *and*

$\Phi_{n+1}(\Phi_n(c_n)) = \Phi_n$ *for all* $n \geq 1$;

Moreover, since $\Phi_0(c_0) = c_1$ we have $\Phi_0 = \Phi_1(c_1)) = c_2$, and inductively,

$\Phi_n = \Phi_{n+1}(\Phi_n(c_n)) = \Phi_{n+1}(c_{n+1}) = c_{n+2}$ $(n \geq 0)$,

in other words, $c_n = \Phi_{n-2}$ *for all* $n \geq 2$, *so that*

$\Phi_{n+1}(\Phi_n(c_n)) = \Phi_{n+1}(c_{n+1}) = \Phi_{n+1}(\Phi_{n-1}) = \Phi_n$,

where the morphisms Φ_n play alternatively the role of argument, function and value. We have then three different but equivalent ways to state the recursive relationship between the Φ_n:

1. $\Phi_{n+1}(\Phi_n(c_n)) = \Phi_n$

2. $\Phi_{n+1}(\Phi_{n-1}) = \Phi_n$

3. $\Phi_{n+1}(c_{n+1}) = \Phi_n$

Proof. The proposition follows easily applying recursively the hypothesis on $\mathcal{C}$ □

Remark 3. *Notice that equality 3. in the proposition above may be written* $ev_{c_{n+1}}(\Phi_{n+1}) = \Phi_n$ *in terms of the "evaluation at x" mappings $ev_x : f \mapsto f(x)$. So the following* projective (or inverse) system *of mappings and elements arises, where each C_n "projects" onto C_{n-1} :*

$$C_1 \overset{ev_{c_0}}{\leftarrow} C_2 \overset{ev_{c_1}}{\leftarrow} C_3 \overset{ev_{c_2}}{\leftarrow} .. \overset{ev_{c_{n-2}}}{\leftarrow} C_n \overset{ev_{c_{n-1}}}{\leftarrow} \dots$$

$$c_1 \overset{ev_{c_0}}{\leftarrow} \Phi_0 \overset{ev_{c_1}}{\leftarrow} \Phi_1 \overset{ev_{c_2}}{\leftarrow} \dots \overset{ev_{n-2}}{\leftarrow} \Phi_{n-2} \overset{ev_{c_{n-1}}}{\leftarrow} \dots$$

Recall that in the category of sets, every such system of mappings,
$C_1 \xleftarrow{p_1} C_2 \xleftarrow{p_2} C_3 \xleftarrow{p_3} ... \xleftarrow{p_{n-1}} C_n \xleftarrow{p_n} C_{n+1} \xleftarrow{p_{n+1}} ...$
has a (projective) "limit", which is rigorously defined as the set C^∞ consisting of all sequences $(c_1, c_2, ..., c_n, ...)$ of "coherent" choices of elements $c_n \in C_n$ ("coherent" meaning here that each c_n "projects" onto c_{n-1}, i.e. $p_{n-1}(c_n) = c_{n-1}$). This projective limit set C^∞ "projects" also in a natural way onto each C_n, sending each sequence to its $n-$th term c_n. Intuitively, this construction allows us to get hold as elements in the limit set C^∞, of "mythical" or "ideal" objects" that cast a series of approximating down to earth "shadows" (the c_n's).

For instance, fractals, a paradigmatic example of "mythical shapes", may be looked upon in this way, as projective limits of everyday shapes (Soto-Andrade and Varela, 1984). In most categories of structured sets the set theoretical limit set C_∞ of the C_n's is also an object in the category, projecting itself by morphisms onto each C_n.

Theorem 4. *Assume that the replacement category $\mathcal{C}$ has projective limits. Keeping the notations of Proposition 2, denote by $\Phi_\infty = \lim_{n\to\infty} \Phi_n$ the coherent sequence Φ_n in the system of evaluation maps ev_{c_n}.*

Then we have that Φ_∞ is a solution to Ouroboros equation: $\Phi_\infty(\Phi_\infty) = \Phi_\infty$.

Proof. Notice that the coherent sequence Φ_n in the system of evaluation maps ev_{c_n} *is* an element of the projective limit C^∞. We may check then that by passing to the limit as n tends to ∞ in the recursive relation $\Phi_{n+1}(\Phi_{n-1}) = \Phi_n$, we indeed obtain the stunning self referential equation $\Phi_\infty(\Phi_\infty) = \Phi_\infty$. □

Apparently no mathematician imagined this recursive procedure to construct solutions of Ouroboros equation before Rosen introduced his $A \xrightarrow{f} B \xrightarrow{\Phi} H(A, B)$ setup as a somewhat opaque formal description of metabolism. This shows how a rather weird idea from biology could show the way to some unexpected and interesting mathematical phenomena. In general the field of biomathematics has been a consumer of mathematical thinking (for example using sophisticated tools to solve ordinary differential equations) but has never given raise to a line of research inside mathematics. We think that this idea of infinite regress is a clear case of a biological issue triggering some deep mathematical questions.

2.3. **Example: A baby arithmetical avatar of Ouroboros.** (Soto-Andrade et al., 2011)

In the category of finite abelian groups, we put $C_0 = C_1 = A = \mathbb{Z}_m^+$, the set of integers mod m, endowed with addition mod m. Then $C_2 = H(A, A) = \{h_a | a \in A\} \simeq A$, where $h_a : b \mapsto ab$ for all $b \in A$ and we see recursively that $C_n \simeq A$ for all n. To identify the mappings Φ_n we need then only to solve multiplicative equations $ax = b$ mod m in A.

If $m = 3$, for instance, we choose $c_0 = 1$ mod 3 and $\Phi_0 = h_2 = 2$. Then $c_1 = 2$ and $\Phi_1 = h_1 = 1$, and our coherent sequence begins $1 \xleftarrow{h_2} 2 \xleftarrow{h_1} 2$. Next, we must look for Φ_2 such that $\Phi_2(2) = h_1$, i.e. for $a \in A$ such that $a \cdot 2 = 1$, so $a = 2$.

It follows recursively that our sequence will look like

$$1 \xleftarrow{h_2} 2 \xleftarrow{h_1} 2 \xleftarrow{h_2} 1 \xleftarrow{h_2} 2 \xleftarrow{h_1} 2 \xleftarrow{h_2} 1 \xleftarrow{h_2} 2 \xleftarrow{h_1} 2 \xleftarrow{h_2} \dots$$

so, intuitively, Φ_∞ is the "limit" of this "wave like" oscillating sequence, although formally Φ_∞ *is* this sequence. Notice also that our sequence Φ_∞ is a multiplicative analogue mod 3 of the ubiquitous Fibonacci sequence! If we take now $m = 10$, and we put $c_0 = 3$ and $\Phi_0 = h_9$, so that $c_1 = 7$, we find recursively that Φ_∞ is embodied in the projective sequence

$$3 \overset{h_9}{\leftarrow} 7 \overset{h_3}{\leftarrow} 9 \overset{h_7}{\leftarrow} 7 \overset{h_9}{\leftarrow} 3 \overset{h_7}{\leftarrow} 9 \overset{h_3}{\leftarrow} 3 \overset{h_9}{\leftarrow} 7 \overset{h_3}{\leftarrow} 9 \overset{h_7}{\leftarrow} 7 \overset{h_9}{\leftarrow} 3 \ldots$$

Translating back into Rosen's original terminology, we have here $a = 3,\ b = 7,\ f = 9,\ \Phi = 7$, but $\beta = (ev_b)^{-1} = 3$, the inverse of b. So β may be identified with b^{-1} but not with b, as pointed out in Cárdenas et al. (2010). Notice that Rosen's demanding assumption on the invertibility of the evaluation at b $(= c_1)$ is satisfied in this baby arithmetical realization, where in fact *all* evaluation maps are invertible.

3. Ouroboros in metabolic systems and categories

Biologists adhering to the new trend called Systems Biology are exploiting optimization principles and advanced simulations based on ODE. We think otherwise: all these analytical tools should be used in conjunction of advanced algebraic tools like the ones derived from category theory.

Our approach to Ouroboros equation here is a further elaboration of the ideas in Jaramillo et al. (2010), where a framework for treating molecules as operators was proposed, in an attempt to relate the theories of *(M,R)* systems and Replicative Autocatalytic Sets (Hordijk and Steel, 2004), We will use here the term "metabolism" as synonym of "metabolic network".

3.1. Protometabolic and metabolic graphs and categories.

We introduce now the notion of a *protometabolic graph* $\mathcal{G}_M$, which is a directed graph with an extra structure, where each oriented edge (also called arrow) appears as having a multi source and a multi target, consisting of sets of nodes and we are also given a map C assigning to each metabolite $m \in \mathcal{M}$ the fuzzy set of arrows it "catalyzes".

More precisely:

Definition 5. *A protometabolic graph is a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{D}, S, T)$ where $\mathcal{N}$ stands for the set of nodes (the reactants), $\mathcal{D}$ stands for the set of directed edges (the reactions) and S and T are the usual source and target mappings $S, T : \mathcal{D} \longrightarrow \mathcal{N}$, with an extra structure given as follows:*

(1) the set of nodes $\mathcal{N}$ is in fact a subset of the power set $\mathfrak{P}(\mathcal{M})$, of a pre given set $\mathcal{M}$ (thought as consisting of the intervening metabolites); so we will write more explicitly $\mathcal{G}_{\mathcal{M}}$ for $\mathcal{G}$.

(2) a mapping

$$C : \mathcal{M} \longrightarrow \mathfrak{F}(\mathcal{D})$$
$$C : m \longmapsto C(m)$$

where $C(m)$ is a fuzzy subset of D (the reactions that m eventually catalyzes) and $\mathfrak{F}(\mathcal{D})$ stands for the collection of all fuzzy subsets of D.

We will say that a metabolite z appears *in $Im(S)$, resp. in $Im(T)$, if $m \in A$ for some $A \in Im(L)$, resp. if $m \in B$ for some $B \in Im(R)$.*

Moreover the following assumptions are made regarding the catalysis map C. Let us denote by $P(m, f)$ the probability that metabolite m catalyzes reaction f. Notice that $P(m, f)$ is just the probabilistic membership weight of f in the fuzzy set $C(m)$ Then:

PM1 *If $C(m) \neq \varnothing$ then m appears in $Im(T)$*
(catalysts are produced by the metabolism)

PM2 *$P(m, f) = 0$ for all m appearing in $S(f)$ or in $T(f)$*
(no reaction is auto catalytic).

Our viewpoint is to visualize $\mathcal{G}_{\mathcal{M}}$ as a sort of generalized graph whose arrows have multiple sources and targets, which are sets of metabolites. See Figure 2 (left diagram) below for the sort of visualization that we favor (applied to our baby metabolism of subsection 3.3.4), instead of the classical biochemical the notation $A + B \longrightarrow C + D$,. for instance, which is not felicitous for categorists since

it suggests that we could just take $A + B$, direct sum assumed to exist, as source node. However, recalling the universal property of the direct sum this would entail the existence of partial mappings from A to $C + D$ and from B to $C + D$, something that does not really make sense in our biochemical context!

Recall on the other hand, that in the dawn of times catalysts were not so specific and they were capable of helping the time course of many reactions; in these past times the name of the game was indeed moon lightning. Since the original first living systems (about 3800 million years ago) a progressive tuning process has shaped the amazing specificity (and efficiency) of modern biological catalysts (enzymes). So our construction allows for metabolic graphs, or metabolic categories, which are evolving entities. Concretely, as time unfolds, the probabilities $P(m, f)$ (i.e. the probability that a given metabolite m catalyses a given reaction f) change. This is one way to model the aforementioned process of tuning that the network undergoes as the catalytic elements becomes more specific in their facilitating action. In this sense our construction shares some resemblance with the *Systems Evolutifs avec Mémoire* introduced in the 80s by Andrée Ehresmann, who had since claimed that a correct representation of living systems can be achieved with the help of time-varying Categories (Ehresmann and Vanbremeersch, 1986).

Definition 6. *We will say that a protometabolic graph $\mathcal{G}_M$ is a metabolic graph iff $\mathcal{G}_M$ is simple (aka minimal), i. e. it admits no proper protometabolic subgraph.*

Definition 7. *We say that the protometabolic graph $\mathcal{G}_M$ is generic, iff it has no inner symmetry (automorphism) other than the identity.*

So in a nutshell, a proto-metabolic graph is essentially a (generalized) graph whose arrows have multiple sources and targets, with an extra datum (the map C) which tells us which nodes (metabolites) act eventually as arrows (= transformations of nodes).

Definition 8. *The (proto) metabolic path category $\mathcal{C}_\mathcal{M}$ generated by a (proto) metabolic graph $\mathcal{G}_\mathcal{M}$ may be defined in the following way:*

- *Its set of objects is the set $\mathcal{N}$ of nodes of $\mathcal{G}_\mathcal{M}$*
- *Its set of arrows is the set of chains $(f_1, \cdots, f_k)$ of arrows in $\mathcal{D}$ such that $S(f_{i+1}) \cap T(f_i) \neq \varnothing$ for $1 \leq i \leq k = 1$, with composition given by concatenation and the identity arrows given by the trivial reactions $n \longrightarrow n$ $n \in \mathcal{N}$.*

If $\mathcal{G}_M$ is a metabolic graph, we say that the associated path category $\mathcal{C}_M$ is a metabolic category.

The way we visualize the composition in this associated category is exemplified in Figure 2 (left diagram) below, for the case of the arrows $\{S, T\} \longrightarrow ST$ (catalyzed by STU) and $\{ST, U\}$ (catalyzed by SU).

3.2. **Enacting (proto)metabolic graphs and categories.** Interestingly, a protometabolic graph $\mathcal{G}_M$, or the associated protometabolic category $\mathcal{C}_M$, can be *enacted*, by choosing a subset $\mathcal{N}_0$ of its set $\mathcal{N}$ of nodes and letting the category recursively act on itself (i. e. "unleashing" the category), so that its arrows act feeding on the nodes in $\mathcal{N}_0$. Metaphorically, we are seeing the arrows as IF A and B, THEN C statements. In this way a "time evolution" or "forward orbit " of the system arises, starting from the "initial state" $\mathcal{N}_0$, as a dynamical system with discrete time, which will tend either to a stable state or to a sempiternal cycle. See examples in section 3.3 below.

The states of this associated dynamical system, associated to $\mathcal{C}_M$, with initial state $\mathcal{N}_0$ are just the subsets of $\mathcal{N}$ obtained recursively from $\mathcal{N}_0$ after all possible reactions are simultaneously enacted,

collecting all their RHS (targets). "Possible reaction" means here that the source and the catalyst for the corresponding arrow appear in the previous state.

Definition 9. *We will call a protometabolic graph or category* primitive *in case the "whole state" $\mathcal{N}$ cannot be reached through enaction from any proper subset state $\mathcal{N}' \subset N$.*

We agree to to code as reactions of the form $1 \xrightarrow{1} m$, catalyzed by 1 itself, all metabolites m in $\mathcal{M}$ not produced by the metabolism (which are available from the environment, like "maná"), where 1 stands for the empty set $\varnothing \in \mathfrak{P}(M)$, the initial object in the category of sets. Biochemically, the object 1 stands for the environment seen as a virtual substrate and for the empty molecule $\varnothing$ as catalyst. So we may assume that every metabolite $m \in M$ appears as the target of some arrow in $\mathcal{D}$.

Remark 10. *There is a natural notion of order among metabolites in $m \in \mathcal{M}$, defined as the minimum number of non trivial reactions which concatenate to produce m. For instance, if $1 \xrightarrow{1} m$, then m is of order 0, i.e. metabolites available from the milieu are of order 0. Typically enzymes are high order metabolites. For example, in Example 3.3.4 below, metabolites S, T, U are of order 0, metabolites SU, ST are of order 1, and metabolite STU is of order 2.*

The underlying idea is that (proto)metabolic graphs or categories are naturally able to apply to themselves. In fact, they would fully deserve their name if they are self-reproducing, i. e. they are solutions of the Ouroboros equation.

3.3. **Baby examples of (proto)metabolic graphs.**

3.3.1. *Example 1*

. $$1 \xrightarrow{1} a \quad 1 \xrightarrow{1} d \quad a \xrightarrow{c} b \quad d \xrightarrow{b} c$$

Enaction of this protometabolic graph starting with $\{a, d, b\}$ ends up in a loop $\{a, d, b\} \rightleftarrows \{a, d, c\}$ and the same starting with $\{a, d, c\}$, so the graph is primitive. It is also simple, but not generic (it admits a flip) and its C is unique. It can be visualized as illustrated in Figure 1 below, as the following baby examples.

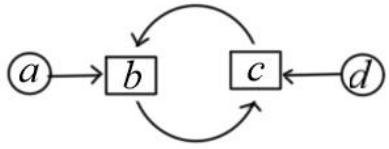

FIGURE 1. Diagram for Example 1

3.3.2. *Example 2*

. $$1 \xrightarrow{1} a \quad 1 \xrightarrow{1} d \quad a \xrightarrow{c} b \quad d \xrightarrow{e} c \quad c \xrightarrow{b} e$$

This protometabolic graph is simple and primitive.

3.3.3. *Example 3*

. $$1 \xrightarrow{1} a \quad 1 \xrightarrow{1} d \quad 1 \xrightarrow{1} f \quad a \xrightarrow{c} b \quad d \xrightarrow{e} c \quad f \xrightarrow{b} e$$

Notice that this protometabolic graph is simple and primitive but not generic.

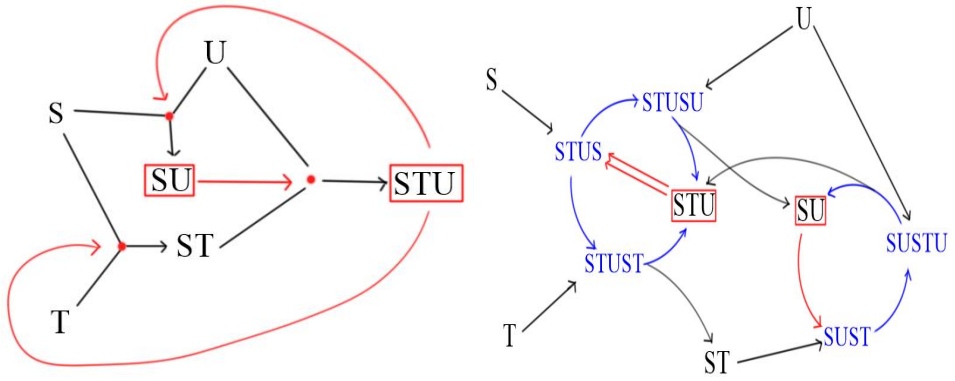

FIGURE 2. Protometabolic graph of a simple reaction network. Left diagram: A system composed of three reactions: S + T $\xrightarrow{STU}$ ST, S + U $\xrightarrow{STU}$ SU,ST + U $\xrightarrow{SU}$ STU. The catalytic role of SU(for one reaction) and STU (for two reactions) is indicated by the red arrows. Please note that *all* the catalysts are produced by the network itself. Right diagram: A more complete (zoomed in) version of the network showing the elementary mechanisms behind each reaction. This example shows how every catalytic event, when explored at higher resolution, can be interpreted as a small network by itself. Thus these two diagrams show different granularity for the formal analysis of the underlying biological network.

3.3.4. *Example 4: The STU metabolic network*

. A more complex example, with more insights from biochemistry is the STU system introduced by Morán et al. (1996) and us (Letelier et al., 2006) intending to explaining the often opaque Rosen papers. The so called STU metabolic network is given by the next three coupled catalyzed reactions, without specifying for the time being the nature of catalysts M_1, M_2, M_3:

$$r_1 : \quad \{s,t\} \xrightarrow{M_1} \{st\}$$
$$r_2 : \quad \{s,u\} \xrightarrow{M_2} \{su\}$$
$$r_3 : \quad \{st,u\} \xrightarrow{M_3} \{stu\}$$

Here Rosen's $f(a) = b$ formalism reads $f((s,t),(s,u),(st,u)) = (st, su, stu)$, In our graph - categorical setting we have : a protometabolic graph $\mathcal{G}_M$ and the associated path category $\mathcal{C}_M$ defined by $\mathcal{N} = \{s,t,u,su,st,stu\}$, the "maná" reactions $1 \xrightarrow{1} s$; $1 \xrightarrow{1} t$; $1 \xrightarrow{1} u$; plus the three reactions r_1, r_2, r_3 above, whose catalysts, given by C, read M_1, M_2, M_3, hitherto undefined.

Many assignments are possible but several such assignments are excluded if we avoid self-catalysis, such as $M_1 = st$ or $M_3 = stu$. This reasoning decreases the initial 27 possible assignments to only 4 (in Rosen's terminology, these would be denoted Φ_i for $i = 1, 2, 3, 4$), to wit:

$C_1 : (M_1, M_2, M_3) = (stu, stu, su)$
$C_2 : (M_1, M_2, M_3) = (stu, st, su)$
$C_3 : (M_1, M_2, M_3) = (su, stu, su)$
$C_4 : (M_1, M_2, M_3) = (su, st, su)$

Each of these four choices C_i for C generates a different protometabolic graph, call it $\mathcal{G}_i$, whose sets of arrows and S and T maps are equal, given as follows (we omit here the obvious "maná" reactions providing s, t, u):

$$\mathcal{G}_1: \begin{array}{c} \{s,t\} \xrightarrow{stu} st \\ \{s,u\} \xrightarrow{stu} su \\ \{st,u\} \xrightarrow{su} stu \end{array} \qquad \mathcal{G}_2: \begin{array}{c} \{s,t\} \xrightarrow{stu} st \\ \{s,u\} \xrightarrow{st} su \\ \{st,u\} \xrightarrow{su} stu \end{array}$$

$$\mathcal{G}_3: \begin{array}{c} \{s,t\} \xrightarrow{su} st \\ \{s,u\} \xrightarrow{stu} su \\ \{st,u\} \xrightarrow{su} stu \end{array} \qquad \mathcal{G}_4: \begin{array}{c} \{s,t\} \xrightarrow{su} st \\ \{s,u\} \xrightarrow{st} su \\ \{st,u\} \xrightarrow{su} stu \end{array}$$

Among these, C_4 is special, as the third reaction of $\mathcal{G}_4$ does not participate in the network because stu is neither the substrate nor the catalyst of another reaction. We therefore discard C_4. Thus from the 27 choices for C that are theoretically compatible with this simple metabolism we have discarded 24, leaving only three as valid assignments : C_1, C_2, C_3.

Notice that in Rosen's terms this means that his β function is a "multivalued function"given by: $\beta(f) = \{C_1, C_2, C_3\}$

The fact that $\beta(f)$ is not single-valued (as any honest function should be) shows that the condition of invertibility of the ev_b mapping, which is the defining property of (*M,R*) systems with organizational invariance, fails for this simple metabolic network. Thus although this metabolic network is an (*M,R*) system, and also an autocatalytic network, it cannot be construed as an organizationally invariant (*M,R*) system because the rule for assigning Φ starting from f gives more than one result. This example is also interesting as it shows that an autocatalytic set is not necessarily a *(M,R)* system with organizational invariance.

In our terms, the condition that β be a honest univalued function translates as the existence of a unique choice for the map C in our protometabolic graph $\mathcal{G}_M$ given by the reaction arrows ri above. It is likely that the unicity of the map C (all other data remaining fixed) is a sensible requirement on a protometabolic graph to deserve the name of metabolic graph.

Most importantly, we see that protometabolic graphs with the same metabolite sets $\mathcal{N}$, same arrow sets $\mathcal{D}$ and same source and target maps S and T but with different C maps, may indeed be different as metabolic networks.

3.4. **Enacting our baby examples of protometabolic graphs.** If we enact our examples of protometabolic graphs, we see that:

For our baby STU example associated to C_1 it is easily checked that the only enaction leading to a non trivial steady state is the one fed with the whole of $\mathcal{N}$; smaller initial states end up in a sempiternal 2 - cycle or die out to the trivial steady state consisting only of s, t, u. In this sense, our baby example is minimal, i.e. simple.

The same happens for the systems associated to C_2 and C_3.

Moreover it can also be checked that the same initial state has different forward orbits when enacted, in the graphs or categories, associated to C_1, C_2, C_3, proving that these graphs are not isomorphic.

Notice also that the protometabolic graph $\mathcal{G}_4$ admits a proper subgraph, to wit the one defined by the first two reactions. In this sense it is not minimal (simple).

If we enact it starting with $\{s, t, u, su\}$ we find the cyclic evolution

$\{s, t, u, su, \} \longrightarrow \{s, t, u, st\} \longrightarrow \{s, t, u, su, \} \longrightarrow \{s, t, u, st\}...$

However, if we enact it starting with $\{s, t, u, su, st\}$ it ends up at state $\mathcal{N}$.

4. Discussion and open ends

Indeed, $f(f) = f$ is an intriguing equation which abstracts phenomena from many fields. Our own interest in this topic arose from a very basic (and unsolved) question in theoretical biology: "What is a correct theoretical framework to formalize systems that construct themselves?". Metabolism is an outstanding example of this self-fabrication as it reconstitutes, through a network of coupled processes, almost all the molecules participating in the network. This intuition, already put forward by Robert Rosen (1958) and especially by Humberto Maturana and Francisco Varela (1972, 1980), with their notions of (M,R) and Autopoietic systems, is centered on the notion of self-reference. Although computer simulations of metabolism (essentially large systems of coupled ODEs) are important efforts in this direction, we feel that to obtain a workable framework to tackle the problem of self-reference is a central quest of modern biology.

As suggested tentatively in this paper, a diagrammatic categorical approach (rather than a purely algebraic one) is a possible pathway. Although Rosen did use the language of categories in his 1950's papers, we claim that what is needed in this context is an extension of the initial notion of categories consisting of objects and arrows. Thus a *metabolic category*, in order to capture the complexities of the Ouroboros equation $f(f) = f$, should be endowed with some sort of generalised morphisms. Its "arrows" are indeed complex entities, suitably visualized as having multi sources and targets, besides objectal identities, representing natural transformations inside self-fabricating devices, acting on objects, but also on other morphisms and *interconverting between objects and arrows*. In this sense, the arrows in a metabolic category should be far more "promiscuous" and multidimensional than the arrows envisaged by McLane and Eilenberg and McLane in 1945, so maybe the framework of n-categories could be fruitfully applied to understand the core phenomenology of living systems. Perhaps this interconvertibility between arrows and objects, a property apparent in biological systems, but rare in Mathematics proper, could be one of the few examples where Biology could induce deep results in Mathematics. Finally, we should not forget the dimension of time which is usually absent in purely algebraic constructions. There seems to be a wide schism between the language of Dynamical Systems (or Processes) and the static structures implied by Category Theory. We have suggested a tentative bridge over that schism built upon the notion of enaction of a (proto)metabolic category, where we see an initial set of metabolites evolving in the course of time according to the "reaction scheme" given by the arrows of the category, thanks to the object-arrow duality realized in a metabolic graph or category. More generally, perhaps some aspects of time-evolution, via functorial transformations, could be explored in this context. Furthermore another relevant dimension of change is the tuning of the specificity of catalysts, allowed for in our model by the variability of the fuzzy sets of arrows associated to each metabolite. As living systems are historical entities, whose present state must be understood in terms of their particular history, we find that the arrows change their specificity over time.

References

Aczel, P. (1988). *Non-well-founded sets*, volume 14 of *CSLI Lecture Notes*. Stanford University, Center for Study of Language and Information, Stanford, CA.

Cárdenas, M. L., Letelier, J. C., Gutierrez, C., Cornish-Bowden, A., and Soto-Andrade, J. (2010). Closure to efficient causation, computability and artificial life. *J. Theor. Biol.*, 263:79–92.

Ehresmann, A. and Vanbremeersch, J.-P. (1986). Un modèle mathématique pour un système vivant évolutif et hiérarchique fondé sur la théorie des categories. *CR. Acad. Sc. Paris*, 302:475–478.

Hordijk, W. and Steel, M. (2004). Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *J. Theor. Biol.*, 227:451–461.

Jaramillo, S., Honorato-Zimmer, R., Pereira, U., Contreras, D., Reynaert, B., Hernández, V., Soto-Andrade, J., Cárdenas, M., Cornish-Bowden, A., and Letelier, J. (2010). *(M,R)* systems and raf sets: common ideas, tools and projections. In Fellermann, H., Dörr, M., Hanczyc, M., Laursen, L., Maurer, S., Merkle, D., Monnard, P., Stoy, K., and Rasmussen, S., editors, *Artificial Life XII: Proceedings of the Twelfth International Conference on the Synthesis and Simulation of Living Systems*, pages 94–100.

Kampis, G. (1995). Computability, self-reference, and self-amendment. *Communication and Cognition - Artificial Intelligence*, 12:91–109.

Kauffman, L. H. (1987). Self-reference and recursive forms. *J. Social and Biological Structures*, 10:53–72.

Letelier, J., Kuboyama, T., Yasuda, H., Cárdenas, M., and Cornish-Bowden, A. (2005). A self-referential equation, $f(f) = f$, obtained using the theory of (*M,R*) systems: overview and applications. In Anai, H. and Horimoto, K., editors, *Algebraic Biology 2005*, pages 115–126. Universal Academy Press, Tokyo.

Letelier, J. C., Soto-Andrade, J., Guíñez Abarzúa, F., Cornish-Bowden, A., and Cárdenas, M. L. (2006). Organizational invariance and metabolic closure: analysis in terms of (*M,R*) systems. *J. Theor. Biol.*, 238:949–961.

Löfgren, L. (1968). An axiomatic explanation of complete self-reproduction. *Bulletin of Mathematical Biology*, 30:415–425.

Maturana, H. and Varela, F. (1973). *De Máquinas y Seres Vivos*. Editorial Universitaria, Santiago, Chile.

Maturana, H. and Varela, F. (1980). *Autopoiesis and Cognition: the Realisation of the Living*. D. Reidel Publishing Company, Dordrecht, The Netherlands.

Milner, R. (2006). *Turing, Computing, and Communication*, pages 1–8. Springer, Berlin.

Rosen, R. (1958). A relational theory of biological systems. *Bull. Math. Biophys.*, 20:245–260.

Rosen, R. (1959). On a logical paradox implicit in the notion of a self-reproducing automaton. *Bull. Math. Biophys.*, 21:387–394.

Rosen, R. (1991). *Life Itself*. Columbia University Press, New York.

Scott, D. S. (1972). Continuous lattices. In Lawvere, F. W., editor, *Toposes, algebraic geometry and logic (1971 Dalhousie University Conference)*, volume 74 of *Lecture Notes in Mathematics*, pages 97–136. Springer-Verlag, New York.

Scott, D. S. (1973). Models for various type-free calculi. In Suppes, P., Henkin, L., Joja, A., and Moisil, G. C., editors, *Logic, Methodology and Philosophy of Science IV: Proceedings of the Fourth International Congress*, volume 74 of *Studies in Logic and the Foundations of Mathematics*, pages 157–188. North-Holland, Amsterdam.

Soto-Andrade, J., Jaramillo, S., Gutierrez, C., and Letelier, J. (2011). Ouroboros avatars: a mathematical exploration of self-reference and metabolic closure. In Lenaerts, T., Giacobini, M., Bersini, H., Bourgine, P., Dorigo, M., and Doursat, R., editors, *ECAL2011: Proceedings of the Eleventh European Conference on the Synthesis and Simulation of Living Systems, The MIT Press, Cambridge MA* , pages 763–770.

Soto-Andrade, J. and Varela, F. J. (1984). Self reference and fixed points. *Acta Appl. Mathematica*, 2:1–19.

Varela, F. and Goguen, J. A. (1978). The arithmetic of closure. *Journal of Cybernetics*, 8:291–324.

DEPARTAMENTO DE BIOLOGÍA, FACULTAD DE CIENCIAS, UNIVERSIDAD DE CHILE

DEPARTAMENTO DE MATEMÁTICAS, FACULTAD DE CIENCIAS, UNIVERSIDAD DE CHILE

DEPARTAMENTO DE MATEMÁTICAS, FACULTAD DE CIENCIAS, UNIVERSIDAD DE CHILE