Ouroboros avatars:
A mathematical exploration of Self-reference and Metabolic Closure

Jorge Soto-Andrade[1], Sebastián Jaramillo[2] ,Claudio Gutiérrez[3] and Juan-Carlos Letelier[2]

[1] Departamento de Matemáticas, Facultad de Ciencias, Universidad de Chile, Casilla 653, Santiago, Chile
[2] Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Casilla 653, Santiago, Chile
[3] Departamento de Ciencias de la Computación, Universidad de Chile, Blanco Encalada 2120, Santiago, Chile
sotoandrade@u.uchile.cl

Abstract

One of the most important characteristics observed in metabolic networks is that they produce themselves. This intuition, already advanced by the theories of Autopoiesis and *(M,R)*-systems, can be mathematically framed in a weird looking equation, full of implications and potentialities: $f(f) = f$. This equation (here referred as *Ouroboros equation*), arises in apparently dissimilar contexts, like Robert Rosen's synthetic view of metabolism, hyperset theory and, importantly, untyped lambda calculus. In this paper we survey how Ouroboros equation appeared in those contexts, with emphasis on Rosen's *(M,R)*-systems and Dana Scott's work on reflexive domains, and explore different approaches to construct solutions to it. We envision that the ideas behind this equation, a unique kind of mathematical concept, initially found in biology, would play an important role towards the development of a true systemic theoretical biology.

Introduction

Ouroboros (also written Uroboros), the ancient symbol of the snake eating its own tail, is often taken nowadays to represent self-reference and circularity. In this vein we call in this paper "Ouroboros equation", the ultimate self-referential equation $f(f) = f$.

Notice that 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.

Recall that equation solving in mathematics has a long history, beginning with equations like $2x = 1$, $x + 3 = 1$, up to $x^2 = 2$ and $x^2 = -1$.

Each of these equations was solved introducing new species of numbers, some of them meeting strong resistance, like negative and imaginary numbers. Indeed methods developed to construct the irrational $\sqrt{2}$ and the imaginary $\sqrt{-1}$ may serve as metaphors to tackle the bigger and subtler challenge of constructing somehow solutions of Ouroboros equation $x(x) = x$. Since this equation suggests that x should be some sort of function, we will write it

$$f(f) = f.$$

in the sequel. However the main 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. For these reasons, we call "Ouroboros avatars", the various manifestations or ways in which Ouroboros equation has emerged in different domains (although "avatar" means in fact "descent" in Sanskrit). We have then 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., 2006, 2005), and others, that we review in the next sections.

A most remarkable fact, commented below, is the similarity of methods of constructing solutions to Ouroboros, developed in fields apparently as unrelated as logic (Scott, 1972, 1973) and metabolic systems theory (Letelier et al., 2006, 2005), motivated by the construction of actual mathematical models for untyped lambda calculus and virtual infinite regress in metabolic systems, respectively.

Ouroboros is not an oxymoron

To begin with, it can be proved that Ouroboros is not an oxymoron, 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). It had been argued nevertheless that this was impossible (Wittgenstein, 1961) or paradoxical (Rosen, 1959). Instead, it turns out that an atomically self-reproducing entity can be axiomatized, and in this sense it really does exist (Löfgren, 1968). In fact Löfgren (1968) has shown that the axiom of complete self-reference is independent from usual set theory and logic, and can therefore be added to it as a new primitive axiom, that it is impossible to derive from the other axioms. Solutions to Ouroboros, as Quine's atoms $Q = \{Q\}$ (Quine, 1980), appear then as completely self-referential, inapproachable, a perfectly closed class in itself (Kampis, 1995). Varela takes a similar stance, when he introduces self-referentiality from scratch as a third mark for self-indication or autonomous value (Varela, 1975), extending the indicational calculus of Spencer Brown (1969), and later as a third logical value, besides true and false (Varela, 1979; Kampis, 1995).

Our viewpoint is however that Ouroboros lives indeed outdoors, with respect of our usual logical - mathematical realm, but just outside, in front of the door, say, so that it can be *approximated* stepwise " from within". This intuition 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), as we explain below.

Ouroboros in Self-referential formalisms

As already said, Ouroboros equation $f(f) = f$ involves self-reference, or more precisely, recursion (for a systematic overview of fields that deal with different forms of self-reference see Kauffman (1987)).

An interesting notion of recursion arises when dealing with its operative issues. This approach, linked with the theory of computing, has a strong relationship with the notion of *application*. It is not surprising that formalisms for abstracting the notions of function and program, like lambda calculus and the theory of recursion, are at the center of these developments.

The paradigmatic theory of functional application is the *simple* lambda calculus (i.e. with no distinction of types) (Barendret, 1984), introduced by Church (1951). In the untyped lambda calculus the equation $f(f) = f$ has a trivial solution: $\lambda x.x$, that is, the identity function. The crucial point here is the absence of typing, something that cannot be realized with the identity function in classical mathematical structures (like vector spaces, groups, etc.), where argument and function belong to different types.

The very essence of the power of this formalism resides in that it overcomes the traditional mathematical notion of function as a set of pairs (input, output), by focusing instead on the composition and evaluation of functions. So formalisms like the lambda calculus are much better suited for the formalization of fields where the process of evaluation is most relevant or even the core of the the phenomenon itself. Lambda calculus was disregarded by the logical and mathematical communities until the seventies. What brought their attention to lambda calculus was the work of Dana Scott providing mathematical models for this formalism. The idea is simple (not so much its implementation however...): finding spaces where these objects (lambda terms, that is, generalized functions) may live. To see the difficulties, let us exemplify the hierarchy of objects that can be created from a set U: functions with zero parameters (these are the elements of U); functions with one parameter, that is, $f : U \to U$; functions with two parameters, $g : U \times U \to U$ and so on. All of them can be expressed in lambda calculus, that is, they should be elements of the wanted space D. In particular, in this typeless environment it should be possible to apply a function $f : D \to D$ to itself, as another element of D.

It is worth reviewing the basic construction in Scott (1972, 1973), where continuity and limits play a central role, by restricting the universe of functions to be considered. The central question is:

"Are there nontrivial spaces D that can be identified (as topological spaces) with their function spaces $[D \to D]$, consisting of all continuous functions from D to D?"

Scott showed that indeed there are many of them, and called them "reflexive domains". His idea was to start with a space D_0, with suitable properties (e.g. a continuous lattice), and try to identify its function space $D_1 = [D_0 \to D_0]$ with D_0. A difficult task indeed, but we may notice that D_0 can be embedded in D_1, by identifying each element $d_0 \in D_0$ with the constant function in D_1 with value d_0, and also that D_1 can be projected onto D_0 by sending each (continuous) function $d_1 \in D_1$ to its minimum value $d_1(\perp)$ (where $\perp$ is the least element of the complete lattice D_0). Call i_0 and p_0 the embedding and the projection so defined. This allows us to embed in a clever way $D_1 = [D_0 \to D_0]$ into $D_2 = [D_1 \to D_1]$, by sending each d_1 to $i_0 \circ d_1 \circ p_0$ and - dually - to project D_2 onto D_1 by sending d_2 to $p_0 \circ d_2 \circ i_0$ and so on, to obtain iteratively a double chain of embeddings from D_n into $D_{n+1} = [D_n \to D_n]$, and projections from D_{n+1} onto D_n, for all n. We obtain then the wanted reflexive domain as the limiting space D_∞ of this double sequence of continuous maps between continuous lattices.

Regarding our interest here, the later result shows that there is a space where Ourboros equation at least makes sense, i.e it "types". To the best of our knowledge, Scott did not consider this equation explicitly, although several notions of his come close to it.[1]

Scott's construction inspired the limiting construction of a self-referential extension of Spencer Brown (1969) calculus of indications by Varela and Goguen (1978), where they endow the collection of all forms that can be constructed in Brown's setting with the same sort of structure that Scott (1972) considered, i.e. chain complete partially ordered sets (posets). In their setting fully self-referential equations like Ouroboros' would have solutions. That is a different way to extend Brown's setting that the one in Varela (1975).

Scott's construction also inspired later the construction of reflexive domains in the context of posets and monotone mappings with suitable continuity properties, carried out in Soto-Andrade and Varela (1984), where the relationship between the existence of fixed points and several instances of self - reference is also discussed (notice that a reflexive domain D is a fixed point for the function $D \mapsto [D \to D]$).

Another formalism where Ouroboros equation arises naturally is hyperset theory (also called non well founded set theory). Hypersets constitute an extension of usual set theory, that allow sets to be members of themselves, like Quine's atom $Q = \{Q\}$ (Quine, 1980; Aczel, 1988). We

[1] See for example Proposition 3.14 in Scott (1972)

meet among them baby Ouroboros like $f = \{(f, f)\} = \{\{f, \{f\}\}\}$, that satisfy $f(f) = f$, if we identify the function f with its graph and choose the usual set theoretical model $\{a, \{b\}\}$ for ordered pairs (a, b).

As discussed in Löfgren (1968) and Kampis (1995), self-reference is closely tied to language. Hence it is not surprising that formalisms that allow to break the classical hierarchies between language and metalanguage, or as in hypersets, between container and containee, can provide solutions to the Ouroboros equation. Up to now however, these formalisms do not seem to have been meaningfully exploited in the context of biological self-reference and circularity (see Cárdenas et al. (2010) more a more detailed survey).

Ouroboros in *(M,R)* systems: Infinite regress face to face

We turn now to 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 Cárdenas et al. (2010)). 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, that transforms inputs $a \in A$ into outputs $b = f(a) \in B$.

But in any metabolic system, catalysts are subject to degradation, wear and tear, and therefore need to be regenerated or replaced by the system. To meet this requirement, Rosen looked upon the replacement mechanism as a procedure, denoted by Φ, that, from a suitable $b = f(a) \in B$ as input, reproduces f according to $\Phi(b) = f$. Because the net effect of Φ is to select from the relatively large set $H(A, B) \subset Map(A, B)$, of all possible metabolisms, a specific f such that $f(a) = b$, using $b \in B$ as an input, Rosen calls it a *selector*. Thus, the procedure Φ representing replacement appears as a map from B to $H(A, B)$.

Then an *(M,R)* system has the following algebraic description based on two mappings f, Φ acting in synergy:

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

But now, it is possible to go further and demand the system to be capable of replacing the replacer, or selector, Φ: a replicative *(M,R)* system in Rosen's terminology (this property is also referred as *organizational invariance* (Cárdenas et al., 2010)). More precisely, Φ should be generated with the help of a procedure that, given a metabolism f, produces the corresponding Φ that selects metabolism f, that is a mapping $\beta : H(A, B) \longrightarrow H(B, H(A, B))$ such that $\beta(f) = \Phi$, and so on...The big question is then, how can this be, without implying infinite regress?

Rosen's solution to avoid infinite regress, was to posit 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 (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. Rosen was however unable to give concrete examples where this hypothesis was fulfilled.

The operation of an *organizationally invariant (M,R)* system can therefore be viewed as three mappings (f, Φ, β) acting in synergy:

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

where β is the inverse of the "evaluation at b" operator.

Now, if instead of shunning infinite regress, as Rosen did, we look at it "face to face", a recursive construction emerges, whose first step is motivated by the question:

If you have a map $f : A \to B$, can you find a new map $f_1 : B \to C$ such that for a suitable $a \in A$ you have $f_1(f(a)) = f$ or, equivalently $f_1(b) = f; \quad b = f(a)$?

Of course, the answer to this question, taken at face value, when A, B and C are plain (unstructured) sets and f and f_1 are set mappings, is "Obviously, yes", since you have plenty of maps from one set to another which take a prescribed value on a given point. Just take C to be the set $Map(A, B)$ of all mappings from A to B and f_1 to be any mapping from B to C such that $f_1(b) = f$.

However this question becomes more intelligent when stated in a categorical framework, typically when we consider our sets endowed with some sort of structure and have our maps preserve this structure.

Then, if we take our structured sets to be vector spaces, our maps would be linear; if our sets are posets (i.e. partially ordered sets), our maps ought to be monotone (order preserving). If our sets were endowed with a metric, or distance, then our allowed mappings might be continuous, or even "isometric", i. e. "distance - preserving" mappings. Structure preserving mappings are usually called "homomorphisms". For instance, the homomorphisms between vector spaces are linear mappings.

Now we can state the categorical version of our question: In a category (of structured sets and structure preserving mappings, say), given a homomorphism $f : A \to B$, can you find a new homomorphism $f_1 : B \to C$ such that for a suitable $a \in A$ you have $f_1(b) = f$, where $b = f(a)$?

The subtlety now lies in the fact that to carry over our obvious set theoretical solution to the categorical setting, we need to find among all mappings f_1 such that $f_1(b) = f$, one which is well behaved enough to be a *homomorphism* from the structured set B to *another structured set* C. We would be happy then to know that the set $H(A, B)$, consisting of all homomorphisms from A to B, may be endowed with the same (type of) structure than A and B. If it is the case, we would take C to be $H(A, B)$, and we would be all set up to seek a homomorphism f_1 from B to $C = H(A,B)$, which takes the value f at point $b \in B$.

Recall now 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).

It is clear however that in the category of sets, where the existence of such an f_1 is obvious, uniqueness is impossible (unless B is a singleton). Nevertheless, if you change the underlying category (i.e. the stage for the problem) so as to have a category whose sets of homomorphisms $H(X,Y)$ are much smaller than $Map(X,Y)$, i.e. become more and more selective, existence may become less and less obvious and uniqueness may become more and more possible.

We may hope then for the existence of a *turning point* in the choice of our category, at which the sets of homomorphisms $H(X,Y)$ would have the right size so as to have simultaneously existence and uniqueness of our homomorphism f_1. Rosen's dream was that such *turning points* (or better, *turning categories*) exist, where his hypothesis would be fulfilled! They might indeed be dubbed "metabolic categories."

If we look however infinite regress face to face and we do not care about uniqueness, we could continue our construction above forever, in the spirit of Soto-Andrade and Varela (1984) under a mild hypothesis of existence of our homomorphisms f_1, in the framework of a *concrete category* $\mathcal{C}$, i.e a category of structured sets and structure preserving maps (the only ones that we will consider in this article).

Hypothesis 1. *(Existence of "replacing homomorphisms")*

We assume that given any homomorphism $f : A \to B$ in our concrete category $\mathcal{C}$, we can choose $a \in A$ such that the following hold:

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

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

Notice that this hypothesis requires implicitly that, A and B being any objects in $\mathcal{C}$, the set of homomorphisms $H(A,B)$ should also be an object in $\mathcal{C}$, i.e. it can be endowed with the same structure as A and B. Also, simple examples (see below) show that it is not to be expected that *every* $a \in A$ be $f-$generic for a given $f : A \to B$.

Example 1. In the category of (finite dimensional) vector spaces and linear mappings, our hypothesis is clearly fulfilled. Indeed, if f is the null mapping 0, we just take $a = 0$ and $f_1, f_2, ...$ to be 0 all the way. If $f \neq 0$, take a to be any non zero vector in A, such that $f(a) \neq 0$ and then f_1 to be any linear mapping from B to $H(A,B)$ sending $f(a)$ to f, f_2 to be any linear mapping sending f to f_1, and so on. These (non zero!) linear maps exist recursively by the well known elementary "linear extension property" for finite dimensional vector spaces, saying that you can always construct linear mappings from one vector space V to another that take a prescribed value at a given non zero vector in V.

Example 2. In the category of additive groups and addition preserving maps, we take $A = B = \mathbb{Z}_3^+$, the set of integers $0, 1, 2$ mod 3 endowed with the operation $+$ of addition mod 3. Notice that $1 + 1 + 1 = 0$ mod 3. Then $H(A,A) = \{h_a | a \in A\} \simeq A$, where h_a is the "scaling map" with ratio a, that sends b to ab $(b \in A)$, *which we identify with* $a \in A$, writing $h_a = a$. So we identify the mapping h_a with its value a at 1. The set $H(A,A)$ endowed with the operation of addition of mappings is also an additive group, isomorphic to A, and $h_a + h_b = h_{a+b}$ $(a, b \in A)$.

If we take now f to be the null mapping $h_0 = 0$, we see that for any $a \in A$, every $f_1 : A \to H(A,A)$, satisfies $f_1(f(a)) = f$, since $f_1(f(a)) = f_1(0) = 0 = h_0 = f$. Hence any $a \in A$ is h_0 - generic and we may take f_1 to be h_0, h_1 or h_2 (i.e. such that $f_1(1) = h_0$, h_1 or h_2). The choice of f_1 becomes relevant when we go one step further, asking now for a homomorphism $f_2 : H(A,A) \to H(A, H(A,A))$ such that $f_2(f) = f_1$. In a diagram:

$$\begin{array}{ccccccc} A & \xrightarrow{f} & A & \xrightarrow{f_1} & H(A,A) & \xrightarrow{f_2} & H(A,H(A,A)) \\ a & \mapsto & f(a) & \mapsto & f & \mapsto & f_1 \end{array}$$

Indeed, since $f = h_0$, we have that necessarily $f_2(f) = f(0) = 0 = h_0$, so f is f_1-generic only for $f_1 = h_0$, but not for h_1 or h_2. On the other hand, if we begin with $f = h_2$ instead of h_0, then for any non zero $a \in A$, we find a *unique* $f_1 : A \to H(A,A)$ such that $f_1(f(a)) = f$, since the equation amounts to $f_1(2a) = 2$, i.e. $x2a = 2$, i.e. $x = a^{-1}$, if we write $f_1 = h_x$. So every non zero $a \in A$ is $f-$generic in this case but 0 is not, since $f_1(f(0)) = h_0$.

Applying now our hypothesis recursively, we can construct the following infinite sequence of homomorphisms (and objects) in our concrete category $\mathcal{C}$, issued from any homomorphism $C_0 \xrightarrow{\Phi_0} C_1$ in $\mathcal{C}$:

$$\begin{array}{cccccccccccccc} C_0 & \xrightarrow{\Phi_0} & C_1 & \xrightarrow{\Phi_1} & C_2 & \xrightarrow{\Phi_2} & C_3 & \xrightarrow{\Phi_3} & ... & \xrightarrow{\Phi_{n-1}} & C_n & \xrightarrow{\Phi_n} & C_{n+1} & \xrightarrow{\Phi_{n+1}} ... \\ c_0 & \to & c_1 & \to & c_2 & \to & c_3 & \to & ... & \to & c_n & \to & c_{n+1} & \to ... \end{array}$$

satisfying the following:

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

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

$\Phi_1(\Phi_0(c_0)) = \Phi_0$ 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 $n \geq 1$;

Notice that to have consistent notations, we have renamed A to C_0, B to C_1, C to C_2; f to Φ_0, f_1 to Φ_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

$\Phi_{n+1}(\Phi_n(c_n)) = \Phi_{n+1}(c_{n+1}) = \Phi_{n+1}(\Phi_{n-1}) = \Phi_n$, showing how the homomorphisms Φ_n play here 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's:

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

Remark now that the last one 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 "reverse" sequence of mappings and elements emerges, where each C_n "projects" onto C_{n-1} :

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

$$c_1 \xleftarrow{ev_{c_0}} \Phi_0 \xleftarrow{ev_{c_1}} \Phi_1 \xleftarrow{ev_{c_2}} ... \xleftarrow{ev_{n-2}} \Phi_{n-2} \xleftarrow{ev_{c_{n-1}}} ...$$

This sequence of evaluation maps ev_{c_n} forms what mathematicians call a *projective (or inverse) system of mappings*. In the category of sets and mappings, every such system of mappings, call it

$$C_1 \xleftarrow{p_1} C_2 \xleftarrow{p_2} C_3 \xleftarrow{p_3} ... \xleftarrow{p_{n-1}} C_n \xleftarrow{p_n} C_{n+1} \xleftarrow{p_{n+1}} ...$$

has a (projective) "limit", which is rigorously characterized 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). In concrete categories we may expect moreover that the structure we have on all C_n's will carry over to the limit set C_∞, which will become then a bona fide object in our category, projecting itself by homomorphisms onto each C_n.

Disgression: A baby projective limit. To convey a better insight into projective limits, we recall here a baby example from Soto-Andrade and Varela (1984), that highlights their elementary set theoretical nature.

Consider the increasing nested sequence of finite sets

$$C_n = \{1, 2, ..., n\} \qquad (n = 1, 2, 3, ...),$$

whose union is the set $\mathbb{N}$ of all natural numbers. This sequence of sets becomes a projective system if we "project downwards", or "contract inwards" each C_{n+1} onto the smaller C_n by sending every $m \leq n$ to itself and $n+1$ to n. Call these projections (or contractions) p_n. So on C_{n+1} we have $p_n(n) = n = p_n(n+1)$. The projective limit C^∞ can be intuited now as the set of all numbers in $\mathbb{N}$ plus an extra "mythical boundary point" $+\infty$, situated at the far right of all natural numbers.

Indeed, going back to the precise definition of C^∞, we see that the points $m \in \mathbb{N}$ appear as "limits" of the sequences of coherent choices $(1, 2, ..., m-1, m, ..., m, ...)$ that after a while "stutter" indefinitely or become "constant". But we also have the coherent chain of choices given by $1 \in C_1,\ 2 \in C_2,\ 3 \in C_3,$ and so on. Notice that each $m \in C_n$ is the "ancestor" of the preceding $m-1 \in C_{m-1}$. This sequence of choices represents then our "mythical far right boundary point" $+\infty$, whose $n-$th projection is n. Analogously, we may obtain $\{-\infty\} \cup \mathbb{Z} \cup \{+\infty\}$ as a projective limit. This shows concretely how the projective limit allows us to get hold of "mythical" or "ideal" objects that cast a a series of approximating down to earth "shadows".

Recall that also fractals, a paradigmatic example of "mythical shapes", may be looked upon in this way, as projective limits of everyday shapes (*loc. cit.*).

Properties of the limit objects C^∞ and Φ_∞. The coherent sequence Φ_n in the system of evaluation maps ev_{c_n} *is* an element of the projective limit C^∞. We call it Φ_∞ and we write $\Phi_\infty = \lim_{n\to\infty} \Phi_n$ to convey the intuition that Φ_∞ is a kind of "limit" of the Φ_n's as n tends to ∞. Notice that this quite analogous to the way in which a "rational" person constructs $\sqrt{2}$ with the help of Cauchy sequences of rational numbers. Now, intuitively, by passing to the limit as n tends to ∞ in the recursive relation $\Phi_{n+1}(\Phi_{n-1}) = \Phi_n$ we obtain the stunning self referential equation

$$\Phi_\infty(\Phi_\infty) = \Phi_\infty,$$

saying that Φ_∞ is *a solution to Ouroboros equation*!

Analogously, making n tend to ∞ in the equation $C_{n+1} = H(C_{n-1}, C_n)$, we get

$$C_\infty = H(C_\infty, C_\infty),$$

so that C_∞ is a reflexive domain, as in Soto-Andrade and Varela (1984). We will not go here into the rigorous justification of this passage to the limit, since it involves a more precise description of Φ_∞ as a mapping in $H(C_\infty, C_\infty)$, taking into account the double system of mappings $\Phi_n : C_n \to C_{n+1}$ and $ev_{c_n} : C_n \leftarrow C_{n+1}$, as in Scott (1972).

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 formal description of metabolism (Rosen, 1958; Letelier et al., 2005) . Notice that this construction is quite different although formally analogous to Scott's (Scott, 1972, 1973).

An arithmetical avatar of Ouroboros. Generalizing example 2 above, we put $C_0 = C_1 = A = \mathbb{Z}_m^+$, the set of integers $0, 1, 2, \ldots m-1$ mod m, endowed with the operation $+$ of 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 identify as before each h_a with a*. We endow $H(A, A)$ with the operation of addition of mappings.

Now, since recursively $H(A, A) \simeq A$,

$$H(A, H(A, A)) \simeq H(A, A) \simeq A,$$

$$H(H(A, A), H(A, H(A, A))) \simeq H(A, A) \simeq A$$

and so on, we have that all C_n are isomorphic to A.

To identify the mappings Φ_n we need then only to solve multiplicative equations $ax = b$ mod m in A. If $m = 3$, as in example 2 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
$1 \overset{h_2}{\leftarrow} 2 \overset{h_1}{\leftarrow} 2 \overset{h_2}{\leftarrow} 1 \overset{h_2}{\leftarrow} 2 \overset{h_1}{\leftarrow} 2 \overset{h_2}{\leftarrow} 1 \overset{h_2}{\leftarrow} 2 \overset{h_1}{\leftarrow} 2 \overset{h_2}{\leftarrow} \ldots$

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: Instead of $c_{n+1} = c_n + c_{n-1}$ we have $c_{n+1} = c_n \cdot c_{n-1}$ mod 3.

If we take now $m = 10$, for instance, 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 reasonably identified with b^{-1} but not with b, as pointed out in Cárdenas et al. (2010).

A linear avatar of Ouroboros. We sketch here a linear example where sets of "metabolites" are vector spaces instead of integers modulo m, so that structure preserving mappings are linear. We denote by $M_{m,n}$ the set of all real matrices with m rows and n columns, identified as usual with linear mappings from $\mathbb{R}^n$ to $\mathbb{R}^m$. We put

$$C_0 = \mathbb{R}^2 = M_{2,1},\ c_0 = \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right),\quad C_1 = \mathbb{R} = M_{1,1}$$

and $\Phi_0 = \left(\begin{smallmatrix}1 & 0\end{smallmatrix}\right)$ (the first projection of $\mathbb{R}^2$ onto $\mathbb{R}^1$). Then we find recursively

$c_1 = 1; C_2 = H(C_0, C_1) = M_{1,2} \simeq \mathbb{R}^2;\ c_2 = \Phi_0 = \left(\begin{smallmatrix}1 & 0\end{smallmatrix}\right)$;
$\Phi_1 = \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$, since $\Phi_1(c_1) = \Phi_0$;
$C_3 = H(C_1, C_2) = M_{2,1} \simeq \mathbb{R}^2; c_3 = \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right); \Phi_2 = Id_2 \in M_{2,2}$ or any matrix with first column $\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$;
$C_4 = H(C_2, C_3) = M_{2,2} \simeq \mathbb{R}^4;\ c_4 = \Phi_2,\ \Phi_3$ being any matrix with first column $\begin{pmatrix}1\\0\\0\\1\end{pmatrix}$ if $\Phi_2 = Id_2$;
$C_5 = H(C_3, C_4) = M_{4,2} \simeq \mathbb{R}^8$, and so on, where we identify matrices with row or column vectors reading their coefficients as usual text. Notice the recursive multiplicative Fibonacci rule $d_{n+1} = d_n \cdot d_{n-1}$ for $d_n = dim\ C_n$.

Notice that Rosen's demanding assumption on the invertibility of the evaluation at $b (= c_1)$ is satisfied in the arithmetical realization above, where in fact *all* evaluation maps are invertible. In the linear example, the map ev_{c_1} is still invertible, although the subsequent evaluation maps are not. In particular, any 2×2 matrix with first column $\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$ would do as Φ_2.

Ouroboros in Autocatalytic Sets

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

We look upon a metabolism as a directed graph $\mathcal{M}$ whose set of nodes $P(X)$ is the collection of all subsets of the set X of all metabolites and catalysts involved in the metabolism and whose set of arrows R is given by the reactions $A \to B$ in the metabolism $(A, B \subset X)$. Molecules x in X not produced by the metabolism are coded as reactions of the form $\emptyset \to x$, where the empty set symbol $\emptyset$ stands for the environment seen as a virtual molecule. We assume further that every metabolite $x \in X$ appears in the target of some arrow in $\mathcal{M}$. Catalysts are defined by a map $C : R \to X$ that assigns a molecular identity to the catalyst of each reaction in R. Of course, we assign the empty catalyst $\emptyset$ to any arrow (reaction) with source $\emptyset$.

A *premetabolism* $\mathcal{M}'$ of the metabolism $\mathcal{M}$ is generated by a subset $X' \subset X$, by taking $P(X')$ as the set of nodes of $\mathcal{M}'$ and all arrows in $\mathcal{M}$ whose source lies in $P(X')$, as its set of arrows.

There is now a natural sense in which a premetabolism $\mathcal{M}'$ may be applied to itself, giving raise to a new premetabolism noted $\mathcal{M}' \frown \mathcal{M}'$: look at $\mathcal{M}'$ and just carry out every possible reaction indicated by $\mathcal{M}'$; then collect all the resulting metabolites together to form the metabolite set X'' of the premetabolism $\mathcal{M}' \frown \mathcal{M}'$.

Ouroboros avatar in this context reads then

$$\mathcal{M}' \frown \mathcal{M}' = \mathcal{M}'$$

To illustrate this formalism let us introduce a simple molecular system which is an *(M,R)* system and a Replicative Autocatalytic Set, taken from Letelier et al. (2006):

$$\mathrm{S} + \mathrm{T} \xrightarrow{STU} \mathrm{ST}$$
$$\mathrm{S} + \mathrm{U} \xrightarrow{STU} \mathrm{SU}$$
$$\mathrm{ST} + \mathrm{U} \xrightarrow{SU} \mathrm{STU}$$

This defines a metabolism $\mathcal{M}$ based on $X = \{S, T, U, ST, SU, STU\}$, with R and C given by the three reactions above together with $\emptyset \xrightarrow{\emptyset} S,\ \emptyset \xrightarrow{\emptyset} T,\ \emptyset \xrightarrow{\emptyset} U$. Now, writing just X' for a premetabolism $\mathcal{M}'$, we can calculate for instance:

$\{S, T, SU, STU\} \frown \{S, T, SU, STU\} = \{S, T, ST\}$,
$\{S, T, ST\} \frown \{S, T, ST\} = \{S, T\}$,
$\{S, T\} \frown \{S, T\} = \{S, T\}$,

so this premetabolism dies out to a trivial solution of Ouroboros equation (i.e. one whose associated reactions are all of the form $\emptyset \xrightarrow{\emptyset} x$. On the contrary, we have
$\{S, T, U, SU, ST, STU\} \frown \{S, T, U, SU, ST, STU\} =$
$= \{S, T, U, SU, ST, STU\}$,
i.e. $\{S, T, SU, ST, STU\}$ defines a non trivial solution to Ouroboros equation!

Ouroboros in Autopoietic systems

Before concluding we would like to bring in the theory of Autopoiesis, as it has deep connections to the idea of

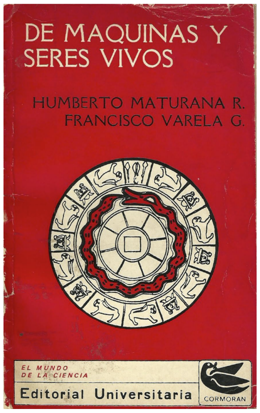

Figure 1: Original cover of the book introducing autopoietic systems (Maturana and Varela, 1973). Although the notion of self-reference is not explicitly mentioned in the book, the authors chose an Ouroboros to illustrate its cover.

self-reference. In fact one of its creators, Varela spent almost a decade looking for a suitable framework to formalize the notions behind this connection (Varela, 1975; Varela and Goguen, 1978; Varela, 1981; Soto-Andrade and Varela, 1984). We won't attempt to reproduce his results, but instead show why self-reference arises from the conceptualization of Autopoiesis theory.

First, we should introduce the perspective of Maturana and Varela for defining a system. A system (or machine) is defined as a unity distinguishable from its surroundings, characterized by two concepts: *organization* and *structure*. The former relates to all processes (or relations) that define the system as a unit and that determine the dynamics of transformations and interactions that the system may undergo as such a unit. The latter are all actual relations that hold between the components of the system in a given space and time (Maturana and Varela (1980), pages 77-84). Now we can define an Autopoietic system (*loc. cit.*) as a network of processes of production, such that its components satisfy the following:

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

ii) constitute the system as a concrete unity in the space in which the components exist by specifying the topological domain of its realization.

The first property of Autopoietic systems can be interpreted as a description of a closed network of production or *metabolic closure*, where the elements needed for the occurrence of each step of the network (such as catalysts) are produced by the network itself. From a dynamical perspective, can also be viewed as non trivial fixed-points of the network dynamics. The notion of *metabolic closure* is common and comparable between several theories of living systems (see Cárdenas et al. (2010) for references). However, Autopoiesis demands more than self-production. What is maintained and reconstituted through the system's dynamics is its *organization*, i.e. what makes it distinguishable as a unit. This is secured in time by the first property and in space by the second. Therefore, if we were to define an autopoietic system we would be tempted to say something like "a unit that regenerates what distinguishes itself as a unit...".

As the last idea suggests, *organizational invariance* can be understood as an ultimate case of recursion or self-reference. In the previous sections we have discussed how to find consistent and non trivial cases where self-reference is possible; future challenges would involve bringing both properties of autopoietic systems into our framework.

Conclusion and final remarks

As we have surveyed here, $f(f) = f$ is an intriguing equation that abstracts phenomena from many fields. It must be underlined that our 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, as the action of metabolism results in the reconstitution of the components that were responsible for its occurrence in the first place.

We are of the opinion that, in order to construct a formalism that captures Metabolism from the perspective of Autopoiesis and *(M,R)*-systems, self-reference is an unavoidable point to consider - not to be confused with simulations of Metabolism, which we regard as complementary efforts. As presented in this paper, dealing with self-reference mathematically, even if it seems to challenge our classical conceptions, is certainly feasible.

Nevertheless, we are aware that the methods exposed are still halfway towards a definitive theory. In particular we should be able to move beyond hypothetical examples into a framework closer to concrete biological systems. Towards this goal there are several avenues for improvement. For instance, so far we have interpreted Metabolism as a network of reactions and catalysis, leaving for later other dimensions of Metabolism, such as time. Self-reference could be regarded as identity conservation under the Metabolic dynamics (Varela, 1975; Varela and Goguen, 1978), and we expect that adding the temporal dimension should allow us to ask more complex questions, closer to molecular systems. Also, we haven't looked closely at the physicochemical properties of Metabolism, which which may provide a grounding as well as a guide for our mathematical models.

In another avenue, one of the main lessons is the vanishing dichotomy between operand and operator, implicit in $f(f)$. This suggest that the phenomena of interaction more than application (in the old functional sense), or concurrency more than sequentiality, may constitute a more appropriate metaphor. As it is well known, life phenomena are intrinsically concurrent, and as such, it appears natural that the emerging formalisms for concurrency are beginning to be applied to this field (Milner, 2009; Cardelli, 2005). We wonder whether there may be avatars of Ouroboros lurking in the concurrent world, an interesting question to explore in future work.

Acknowledgments

We would like to dedicate this paper to the memory and work of Francisco Varela, friend and mentor.

References

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

Barendret, H. (1984). *The Lambda Calculus: Its Syntax and Semantics*. North Holland, Amsterdam.

Cardelli, L. (2005). Abstract machines of systems biology. In Priami, C., Merelli, E., Gonzalez, P., and Omicini, A., editors, *Transactions on Computational Systems Biology III*, volume 3737 of *Lecture Notes in Computer Science*, pages 145–168. Springer, Berlin.

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.

Church, A. (1951). *The Calculi of Lambda Conversion*, volume 6 of *Annals of Math. Studies*. Princeton.

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.

Milner, R. (2009). *The Space and Motion of Communicating Agents*. Cambridge University Press.

Quine, W. V. (1980). *New Foundations for Mathematical Logic*, pages 80–101. Harvard Univ. Press, 2nd edition.

Rosen, R. (1958). The representation of biological systems from the standpoint of the theory of categories. *Bulletin of Mathematical Biology*, 20(4):317–341.

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

Spencer Brown, G. (1969). *The laws of form*. George Allen and Unwin, London.

Varela, F. (1975). A calculus for self-reference. *International Journal of General Systems*, 2(1):5–24.

Varela, F. (1979). The extended calculus of indications interpreted as a three-valued logic. *Notre Dame J.Formal Logic*, 20:141–146.

Varela, F. (1981). Autonomy and autopoiesis. In Gerhard, R. and Schwegler, H., editors, *Self-organizing systems: An interdisciplinary approach*, pages 14–23. Springer-Verlag, New York.

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

Wittgenstein, L. (1961). *Tractatus Logico-Philosophicus*. Routledge and Kegan, London, english edition.