

ORGANISMS AREN'T, THEY HAPPEN

GONÇALO BRAGA

Abstract

This is merely to keep track of where, and how I'm thinking about the problem, in a very informal manner. A sort of log if you will.

Introductory statement

Organisms and life (the process) are actually things that are very illusive. We usually go on to describe proxies of them, as in genetics, molecular biology, and overall life-sciences, but organizational principles and a description of what organismal organization is, is missing. There have been a myriad of attempts at doing so, and a common theme is that of relations. An organism is in virtue of the relations between its constituents. What matters are the relations, and not the constituents themselves. That's why some people (Varela, etc), refer to what we see as an physical instantiation of life. If there would be other substrates that nonetheless show the same relations between themselves, such system would be characterized as an organism.

Even more than this central theme, is one derived from it: self-reference or impredicativity. It means that such systems are self-determined, and can be illustrated very simply, as by what Rosen (presumably one of the first to put the argument into completely relational terms with category theory) pointed out. If we take f as metabolism, we have

$$(1) \qquad f(f) = f$$

Here f is serving as function, argument and result. Mathematics as a whole tries to push away such infinite regress, and such characteristic is the bane of its existence (presumably only under classical logic; such that three-value logics might be able to go around this problem, or even just using constructive logic). A bunch of paradoxes: Liar paradox (Epimenides'), Halting problem, Russel's paradox, etc, derive from this. One needs to understand, though, that under this infinite regress $f(f(f(f(...))) = f$, the ambiguity in understanding where such object is serving under each role (function, argument, etc), is precisely why semantics can't be completely reduced to syntax, and this shouldn't be avoided. It should be actually explored in a full manner, much

like the way that the real domain was expanded into the complex domain by addition of another object $i = \sqrt{-1}$, by having the contradiction $x^2 = -1$. Here x would need to be both positive and negative for it to follow. This is the type of exercise which is taken by three-value logics, and more specifically with Spencer-Brown's calculus of indications, which was later extended by Francisco Varela.

On modelling impredicativity and self-reference

Dynamical systems theory (be it non-linear or not) as far as I'm aware only deals with the evolution of state variables in a deterministic or stochastic manner, according usually to a set of ODEs/PDEs. We might also have more complex behaviour by having some of these being coupled to each other. Over the corresponding phase-space, there can be analysis of the stability of the fixed-points, if they exist, given small perturbations. However what's missing is the notion of blending both operand and operator. In this case, we would have a reflexive space under which any object or state also acts as a transformation. We want the corresponding state-evolution to also affect meta-dynamics (imagining here a changing set of PDEs). Even more important is the stability of organization, that is the stability of relations between constituents. Modelling organismal organization, I assume, needs these types of approaches. On that matter, λ -calculus allows for functions to be both operands and operators. Some concepts in category theory are also useful. Furthermore, there's Spencer-Brown's calculus of indications which does also have some interesting concepts, namely those extended by Francisco Varela, Louie Kauffman, etc.

So in essence, one is looking for a way to express fixed-points over organization, that is, over relationships between processes of the system, and not necessarily fixed-points regarding state-variables. There would be various ways to have the same fixed-point organizationally that nonetheless wouldn't correspond to fixed-points if one is looking at state-variables. This is the problem of modelling impredicativity.

Similar approaches?

Regarding Rosen's approach with category theory, and Varela's with his extension of the calculus of indications, it largely seems (atleast at surface) that these approaches are very similar. Particularly, in the similarity of Varela's third state of re-entry, or the autonomous state, with the concept of an endomorphism. An endomorphism describes an object mapping to itself. The same could be characterized for the autonomous state. Rosen constructed his (M-R) system mapping to avoid the infinite regress of organismal organization (or to express it in a finite form). The same could be said for Varela's ap-

proach, with a three-value logic. One can wonder what would be, if they had actually been aware of each others work. There needs to be a better way of conceptualizing fixed-points over organization, under which there's a reflexive domain. That is a domain, for which the objects also act as transformations.

The ouroboros dilemma

The ouroboros equation is given by $f(f) = f$ as previously mentioned. Although having the correct framework to find non-trivial solutions for such is difficult, I would argue the most difficult part is the association between such object and processes in the natural world. Rosen specifically chooses f as metabolism (as in cellular metabolism), and has other objects and mappings to close his (M, R) -system under which every efficient cause is corroborated by a material one. Such objects and mappings lead to closure to efficient causation, and associated organizational invariance. My question: would this abstract model capture every instantiation of an organism? My contention comes from the "repair" or regeneration part, under which if not the actual components of a set, atleast the corresponding "classes" of such set would need to be regenerated. Why? Why can't we have run-away-like systems for which such sets aren't regenerated if not fully, atleast partially? This comes, I guess from my limited capacity, but of not understanding associated concepts of closure. Even closure of constraints, which is a similar concept, usually could also be said to happen in other dynamical systems which are clearly not alive (weather patterns, fire, etc). So in this capacity one would either affirm that such closure isn't maintained for a long enough time, or it has a small number of associated constraints (which by default would be very unlikely to be regulatory). Which gets us to consider again the obvious: There's a huge amount of evolutionary momentum. Why are we taking some concepts and events for granted? Even more specifically why are we confusing the evolution of such systems, with their ontology. One assumes, typically, that these are related, but this is a very Newtonian assumption. This regeneration of sets, which we typically call metabolism, seems to me to be an assumption. It's clearly easier to maintain closure due to it, but I don't think it's a necessary condition. The same goes for other processes (such as replication, which would fall under regeneration of sets, or a typical boundary), which could be better viewed as adaptations to compensations, in order to maintain a specific organization. This organization, the organismal one, which allows dialectical processes (ones that are co-dependent existentially on each other) to emerge such that they and their relations evolve so as to maintain this very specific organization. In this way this organization Φ is what could be lent to be a fixed-point, and have the associated ouroboros equation $\Phi(\Phi) = \Phi$. The problem is defining such organization in terms of observables that are amenable to inspection (in a practical manner) in both organisms and other dynamical systems.

What I don't understand about (M, R)-systems

I either completely miss what Rosen puts forward, or I actually think they are too specific.

An organizationally invariant (M, R)-system is viewed through the respective mappings:

$$A \xrightarrow{f} B \xrightarrow{\Phi} H(A, B) \xrightarrow{\beta} H(B, H(A, B))$$

for which $f(a) = b$, $\Phi(b) = f$ and $\beta(f) = \Phi$, such that it's closed to efficient causation. The first mapping f refers to metabolism and the corresponding transformation of reactants to products under the action of catalysts, $f : A \rightarrow B$. The second mapping Φ is associated to the "repair" system, which regenerates the corresponding catalysts which allow the morphism f . The corresponding morphism is $\Phi : B \rightarrow H(A, B)$ as it is repairing from all possible sets of metabolisms ($H(A, B)$). Following that, such repair system needs to be generated, and needs to be generated from within. As such we have a third mapping $H(A, B) \xrightarrow{\beta} H(B, H(A, B))$. As such, the only possible metabolism (the current one) is used so as to regenerate all possible metabolisms, that is, $\beta(f) = \Phi$. In order to avoid the infinite regress, Rosen allows $\Phi(b) = f$ to have only one solution. In this way, β needs no further constraints, and is taken as an inverse of the morphism $\Phi(b) = f$. This is a pretty big constraint, as noted by Soto-Andrade et al. (2011). If we interpret this system, we have a certain metabolism f , whose catalysts (which are eventually degraded) are eventually regenerated by Φ , and which repair system is then generated from within with β . However, the only metabolism that is allowed to be used under the morphism β is f and not any other metabolism in the set of all possible metabolisms $H(A, B)$.

I don't see how this isn't too restrictive, and I guess Soto-Andrade et al. (2011) develop on that. My problem resides, again, on why regenerate a set, multiple sets, or even a category of sets? Why? I legitimately don't understand why such regeneration isn't an assumption, and a big one at that, for the nature of organismal organization? Mistaking the ontology for the physiology?

On material impredicativity

If we take into account material impredicativity, and with it Turing machines, the reading head and corresponding rules that govern it, would need to co-create each other, changing accordingly. In this analogy, an organism would be a Turing machine that, additionally to the prior description, also builds its own reading head ("from within"). This steers with self-reference. Majority of the diagonal arguments used to solve "paradoxes" (Cantor's, Russell's, Gödel's, Turing's halting problem, etc) seem to be using classical logic, which allows proof by contradiction, which wouldn't be allowed in constructive logic given that it does not assume the law of the excluded middle. There needs to be

a better way to talk about self-reference and actually take these "paradoxes" for what they are: Non-existent cartesian dualities. In trying to understand organismal organization, what "shouts at us" immediately is self-reference. There's no way to go around this. These systems define themselves from within. The problem resides in formalizing, if possible, such ontology and corresponding evolution of such systems.

What gives?

There's a peculiar similarity between these "problems", that condense into: semantics can't be fully reduced to syntax, as the consistency of these systems can't be fully addressed from within (Tarski's, Cantor's, Gödel's, etc), and the approach that would be needed to address the self-referential organization in organisms. I go back to a quote from Varela's paper on "A calculus for self-reference":

It appears as if different, successively larger levels are connected and intercross at the point where the constituents of the new lower level refer to themselves, where antinomic forms appear, and time sets in.

Self-reference or impredicativity is the name of the game.

(M, R)-system in type-free λ -calculus

As given in Mossio et al. (2009), although with some contentions with assumptions as discussed in Cárdenas et al.(2010), Rosen's canonical (M, R)-system can be put into untyped λ -calculus in the following manner. The corresponding morphisms would be, given that we associate $\beta = B$:

$$(fA) = B$$

$$(\Phi B) = f$$

and

$$(Bf) = \Phi$$

Replacing B and then Φ we get:

$$f = ((fA)f)(fA)$$

We can now introduce $G = \lambda x.((xA)x)xA$ and the Y combinator as $Y = \lambda y.(\lambda x.y(xx))(\lambda x.y(xx))$, from which :

$$G(YG) = YG$$

follows. Introducing $f = YG$, we then state:

$$f = Gf = ((fA)f)(fA)$$

given that we'll have:

$$B = ((YG)A) = YGA$$

and

$$\Phi = (B(YG)) = (YGA)(YG)$$

As such, we have f , Φ and B all defined in terms of A and the Y combinator. There are some arguments in Mossio et al. (2009) asserting that Rosen's (M, R)-system is not really impredicative, and even if it were, such condition wouldn't preclude it from being computable, as in simulation sake. Cárdenas et al. (2010) say that the computability of such diagram isn't valid given that the association $\beta = B$ is not valid, and that the strenght of the proof given by Mossio et al. (2009) is weaker than Rosen's version which is based on Turing's definition of computability.

Is the concept of closure a necessity, a sufficiency, both or neither?

When talking about closure of constraints or organizational closure, usually it is said that closure is achieved when each of the constraints is generating atleast another constraint in the system, such that the system achieves closure from within. There's another problem though, which is: how should one discern if some thing or some process is a constraint. Mossio and Montévil say that such constraint needs to be conserved (or have symmetry) at a certain time scale τ . If this is a good way to go about it, I'm not sure, as literally everything will be considered a constraint at a certain scale under this definition. More importantly, I believe this suffers from the same problem than Rosen's approach does. When we talk about identifying processes or components to be in a certain set or category of sets, we are inviting a problem. Every component or process, apart from being mainly "abstracted" into a single set, is also going to influence (and be explicitly) in every other set. The system needs to be considered as a whole. Decomposition like this tries to mask away the problem. How are we going to prove that a cell might have closure? With our abstract models? Assuming that the modularity of function we see in biochemical circuits is really what's going on in cells? Talking about N state variables and the corresponding phase-spaces, and then taking the very strong assumption "if all else equal" then our analysis of such space, along with attractors, fixed-points and trajectories is valid? We might be fooling ourselves with this approach even. Take a relationship between N state variables and attribute to them a certain behaviour, and assume the rest of the variables that could potentially be considered to describe the system don't influence such space? Do we go about computing sensitivities between state variables ad-infinitum? This is not being lazy enough. And potentially missing a big point.

One of the problems that seem to be associated to closure, is that other dynamical systems also have it. There are arguments that presumably the complexity of the constraints isn't as high as with the one in organisms (emergence of regulative constraints, which after perturbations control the constitutive constraints, but which need an explanation for their emergence), or that perhaps closure isn't as achieved for enough time. I think something that might make it clear, is that the organization of such systems is different. In organisms, if we take Φ to be an organizational closure function (which would be a function of the underlying processes and relations between them), we have:

$$\Phi(\Phi) = \Phi$$

which in other words means that it is a fixed-point. It can be considered as autonomous organizational closure. In other dynamical systems, we would have:

$$\Phi(\Phi) \neq \Phi$$

In organisms, any type of adaptation (e.g. metabolism, replication, aging, etc) is to be taken as a compensation for a perturbation (even for n -th order perturbations which would emerge from within the system) in order to conserve Φ . Although it presumably gets more complicated as scale is increased. This is a very specific type of organization. It is the one under which the dynamics evolve so as to conserve it. It is inherently self-referential. The problem resides in explaining its emergence.

From such equality and inequality seen prior, we can state that in organisms we have an infinite regress of the type $\Phi(\Phi(\dots)) = \Phi$, and in other dynamical systems this isn't the case. After a certain point the organization function Φ is no longer a "fixed-point". Which it would never be from the start, as fixed-points imply infinite regress. This is I presume the main difference. Even if one could point to organizational closure in both organisms and other dynamical systems, the type of organization seen in organisms is a fixed-point. We should be wary of the huge amount of evolutionary momentum that there is, which leads to confuse the ontology of the organisms with their evolution or physiology. These are presumably very different. Other results obviously associated to organisms, are those of Prigogine and other authors, associated with non-equilibrium and stochastic thermodynamics. Again, organisms are far-from-equilibrium systems, they are thermodynamically open systems, but organizationally closed ones.

At what point (how?) does Φ become a fixed-point. Is closure (of constraints) enough to describe such organization? Is it even necessary? This organization is defined in terms of its capability to conserve itself. The processes and their relations, which "compose" (using the word very loosely) the system, evolve in a manner so as to conserve Φ . How shall Φ be described and defined. Is it equivalent to closure (of constraints)? Wouldn't a run-away system (which

never returns to the same initial set (or classes) of components) be capable of such, eventhough it (presumably) doesn't show closure of constraints? It's as if organisms are a universe of their own, when it comes to their organization or operation, eventhough clearly thermodynamically open.

Organism as a universe of its own

I presume what I mean by organization Φ of a system, such that $\Phi(\Phi) = \Phi$ is more, or atleast different from the concept of closure of constraints. As such, an illustration might provide more insight, given that the self-reference can't be taken head on. If we go back to some though by Ernst Mach, regarding inertia (which by the equivalence principle would be equivalent to gravitation) being originated by all masses in the cosmos, and posterior work by Denis Sciama, Dicke, Schrödinger, and other authors, and also taking into account Dirac's Large Number Hypothesis, we get some expressions, such as:

$$\frac{GM}{c^3\tau} \sim 1$$

where G is the universal gravitational constant, M the mass of the universe, c the speed of light and τ the age of the universe, and taking into account:

$$G = c^2 \sum_i \frac{r_i}{m_i} = c^2 \frac{R}{M}$$

where R is the radius of the universe, we get:

$$\frac{R}{c\tau} \sim 1$$

Regardless of the validity of such, what's interesting is the definition of "laws" and "constants" in a purely relational and impredicative manner, to the extent that the unity might represent was being presented previously as the organization Φ . Such would be referred to as meta-dynamics. Such systems, eventhough thermodynamically open, as is the case of organisms, and which follow very clearly physical "law", are organizationally or operationally closed. Φ would be the "universal" property of such system. There would be an analogue of inertia, at the operational or organizational level. Unless there are perturbations to such system, no process or relations between them are going to change. If a perturbation happens, all other processes (and respective relations) collectively compensate such perturbation (potentially originating n -th order perturbations, which would be compensated from within the system). These processes evolve in order to conserve Φ . Any process or relation is defined from within taking into account all other processes "composing" Φ . What needs to be understood is how such processes in this type of system behave. Not how they look from "outside" (following physical law), but how

they look from "within" (meta-level). Any process in such system would behave in complete relation to all others.

This is a hard problem in analogy, on finding analogous meta-dynamics for such organization. Lest we forget the obvious, that all our measurements and models are relational. All of them take into account measurements of the world with relation to something else. What we see in the world is a reflection of our constraints, and no small world (our models and formalization) or combinations of them will ever exhaust the large and open-ended world in front of us.

What's the organismal ontology?

The real problem resides in understanding the ontology of such systems. Their evolution or physiology, can be captured very vaguely by understanding that any new behaviour or adaptation, can be viewed as a compensation to a perturbation in order to conserve Φ . However, how do they emerge? At what point (how?) does operational closedness emerge? Is the physiology even remotely related to the ontology? Similar? Nothing to do with it? How does operational self-reference happen? All of these problems come from infinite regress, corresponding singularities, and the problem of self-reference. Of not being able to address a possible formalization from the outside, of a system that is completely defined from within (that is, that has operational or organizational closedness).

Is the question not even wrong?

Is this prior question correct? It is "sensible" to ask when (how?) does operational closedness emerge? Perhaps one should look where such ontology is no longer. We had:

It appears as if different, successively larger levels are connected and intercross at the point where the constituents of the new lower level refer to themselves, where antinomic forms appear, and time sets in.

Its opposite:

It appears as if different, successively lower levels (top \rightarrow down scale progression) get disconnected and stop being intercrossed at the point where the constituents of the lower levels stop referring to themselves, where antinomic forms disappear, and time ceases.

What's the takeaway? The system which has operational closedness (and being taken as an analogue of a "universe") shrinks in size? Time or its progression ceases or slows down? More importantly, does any of this make sense? How can one put this into "reasonable" concepts?

The usefulness of allometric scaling

As is known there are some tendencies relating metabolic rates, mass, and life span of organisms. Namely, that of Kleiber's law, which has the basal metabolic rate of an organism which can be equated to heat released, ΔQ in Watts or J/s as depending on mass with:

$$\Delta Q \propto m^\alpha$$

with α usually being $3/4$, although there being some contention, and $2/3$ also being "drawable". Life span is also related usually with the relationship $\tau \propto m^{1/4}$.

What should be clear is that these tendencies reflect relations between constraints of such systems, namely with very good arguments from a thermodynamic underpinning, but they don't (or presumably shouldn't) say much about the ontology of such systems. Perhaps one could assume that the organization Φ of such systems would work as a fixed-point regarding relations between such observables, in a way that:

$$\frac{R}{\tau \Delta Q} \sim \Phi$$

where R being the volume or the dimension of the system (which could be taken as mass given that $m \propto R$ in such allometric scaling), τ and ΔQ would evolve in such a way to conserve Φ . This one being a possible relationship out of multiple other ones. The point being that when such doesn't follow the system loses operational closure and ceases. Aging and associated death would be like breakage points for such relationships, however one should note that these would probably appear as compensations early on (evolutionarily) in order to conserve Φ of another system.

How does an organism emerge?

A good chunk of this is based on the autogen model by Deacon. However, assuming that I'm not misinterpreting his view, what he proposes removes, for lack of a better word, the agency and autonomy of such system. Let me digress a bit. One of his vizualization, or physical instantiations of an autogen is that of a catalysis system (or an autocatalytic network) which using some substrate builds itself (more enzymes) and some boundary/capsid system. He obviously notes, that if such is the case, this will be merely self-organization, and what was once a system far-from-equilibrium will slowly become inert, as there are no distabilizing forces, such that the catalysts are "imprisoned" inside the capsid system, with no further transport of substrate at their disposal. He puts forward that such system when building its boundary might increase to a point where it forms two different assemblies (which would be

akin to replication), and furthermore through the selectivity and sensitivity of the container molecules on the boundary, the respective substrate on the outside (along with respective concentrations) might lead to its instability such that there's some possibility for this system to progress into another state, other than the inert one if it is ever achieved.

The problem, I assume and presuming that I'm not misinterpreting his position, is that by relying on external perturbations to remove the system from an inert state, it's actually "stealing" the system's autonomy. From the get go, a system like this I presume couldn't be considered an organism. The alternative, that I assume would have autonomy, is that of a system which additionally to the autocatalysis of the enzyme system, and the boundary one, would also have a secondary product which would be a (partial) disruptor of such boundary/capsid. In this way, the system wouldn't rely on any external perturbations to get out of the inert state. The inert state is the *dead* state. At such point (Deacon's autogen), we are merely talking about self-organization still, such that it could be an analogue of a virus (in a very loose sense, such that it would converge to an inert state without strong enough external perturbations). This continual reconstruction of the boundary would allow selective permeability, which would allow not only the transport of substrate molecules but also of other molecules which would confer variability to such system. This infinite regress, as what was pointed out before regarding fixed-points and the blend between operand/operator along with the junction of antinomic forms, is what an organism *has*. It inherently provides autonomy to such system by maintenance of such infinite regress, by having internal perturbations. There's no reliance on external perturbations, of which there would none of (at least strong enough) in a very homogeneous molecular environment. If such infinite regress stops, the organism is no longer. The only way for it to become organism again, is to have a strong and direct enough external perturbation such that it brings back the autonomous state, which is comically unlikely.

How does a code emerge?

Still going back to the system approached before, where a self-catalytic system generates both itself (further enzymes), a boundary/capsid system and another (partial) disruptor by-product that destabilizes such boundary, with the use of respective substrate, a "code", or essentially a link between semantics and the physical instantiation of such molecules, what one would typically call biosemiotics, might appear in the following manner:

The RNA world hypothesis has some appeal precisely because of the existence of ribozymes, which also have catalytic function, beyond their "coding" one. But this is after the fact. When do certain molecules have semantic information? If you put DNA/RNA or any class of molecules which might be associated to corresponding compression of information based on past expe-

riences of the system, outside such system, they do nothing, they have no semantic information. Precisely because there's no interpreter system. As such, if one considers the system previously described and considers that such self-catalysis is on the basis of incorporation of nucleotides in a sequential manner, with the corresponding folding, majority of the rest of the system (both boundary/capsid and disruptor system) are "hanging" on the kinetics/specificity with regards to such enzyme composition. And you might do all of these (self-catalytic, boundary/capsid and disruptor system) with just one molecule and rely merely on the oscillation of such conformation over time. On this matter, two of the enzymes could form a capsid molecule when they are in reasonable conformation, much as the destabilization of such being induced by the disruptor one.

Going back, the "code" naturally emerges based on the incorporation of substrate and corresponding sequence of the catalytic unit, relying on the availability and diversity of substrate around. As such, this variability will change corresponding affinities and reaction rates, leading to a different interaction between the self-catalytic, boundary and disruptor systems, which are inherently reflecting the environment (substrate availability and diversity).

The important thing is to keep the infinite regress going. If such stops, the system is *dead*.

Organisms don't depend on external perturbations, and yet are a reflection of the external environment.

Organisms are only *alive* when there's this infinite regress, this process of re-construction of boundaries which allows temporal/selective permeability which will induce some variability in the system. Perfect replication (of components/classes) isn't necessary and would presumably be detrimental as it would lead to excessive self-organization driving the system into the inert state, the *dead* state.

On a (pseudo)autonomous system

If one can't go through testing a physical instantiation of such organization, then one can at least simulate it, or try to. That would be under the form of material impredicativity. Better yet, what is referred to here can be captured by the concept mentioned before, of a Turing Machine which builds its own reading head. This would be a purely algorithmical approach, thus we say that such system would pseudo autonomous. Either way, taking the system described before with such organization as being composed by a self-catalytic system which gives rise to itself, a boundary/capsid system, and a disruptor system for such boundary, such that this system doesn't rely on external perturbations, such algorithmical analogue would be that of a perceptron which builds itself. Better yet, such organization would reflect the environment, which here would be the patterns in the corresponding dataset. Hopfield-nets

or associative memory ones, get close to the extent that the weights of the net are instantiated as $W = X \otimes X^\top$, purely on the basis of the patterns, however such network lacks any autonomy to the extent that the energy minimization of the landscape which is done on the inference step is completely detached from any "necessity" of such network to keep on.

To better exemplify what I mean, in a very crude sense, the patterns (external perturbations, E) of the dataset would iteratively combine to form a W matrix which would represent the self-catalytic system, to the extent that some "residues" of the catalytic process would form the boundary/capsid matrix, B , and the disruptor one, D . To deal with convergence problems, to the extent that the network might get into a state of too much self-organization, the *inert* state, the perturbations which get into the system $E - B$, such that B itself is changed on the basis of D (with a subtracting effect). The mentioned prior of combination, would probably rest on a similar approach to that of Fontana and Buss on the construction of a minimal chemistry with λ -calculus, such that molecules are represented by λ -abstractions, with the new form being obtained by β -reduction. Here, the process although similar would need to differ a bit, as the vectors representing the patterns, aren't combinable in the same manner.

Another option would "just" be to choose a few operations to combine such vectors, although the biggest difficulty is on having W generating multiple other matrices in the following manner: $W^\tau \rightarrow \{W^{\tau+1}, B^{\tau+1}, D^{\tau+1}\}$.

A great deal of intuition might also be derived from better understanding what is, properly, adaptation, learning and code formation, in earlier stages of such systems (similar to the autogen model).

How is a system autonomous?

From Deacon's "Incomplete Nature: How Mind Emerged from Matter":

Recall that unlike most living organisms, autogens do not actively maintain non-equilibrium conditions. They are not continually in a dynamic state, but may spend vastly more time as inert structures.

I presume what's missing from the autogen model is the realization of a system which "generates" itself, a boundary, and a disruptor for such boundary, all from within. "Generating" although it might be associated to auto-catalytic function doesn't imply perfect or even remotely closure of *all* constraints. What I mean by this is that, when such a system becomes self-organized with a corresponding boundary, and a catalytic network which can replicate processes (not necessarily class of components, and definitely not individual components), it will inevitably fall into a state which is inert, which is inherently too self-organized. And so, in a way to prevent the reliance on external perturbations, such system needs to generate internal ones, such that it can

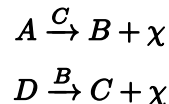
always "steer" away from such inert state. Not lose all self-organizing capacity, but to prevent convergence to the *inert* state. In this manner, any system that is an organism, and which inherently autonomous, is going to keep this type of infinite regress going. That is, the self-organizing capacity along with the permanent re-construction of its boundaries, in such a way that it benefits from the diversity of environmental effectors, but which definitely does not rely on them. As such, these systems eventhough a reflection of the environment, aren't at all dependent on external perturbations, for that would be stripping them of their *agency*.

Organisms aim to maintain their agency, they aim to maintain the infinite regress. Everything else can be seen as compensations in order to keep it that way.

What we see in self-replication and reproduction are merely some physical instantiations, and good ones at that, of replicating some constraints which are crucial to the maintenance of this self-catalytic with a boundary and a disruptor system.

On Deacon's autogen and the offloading of constraints onto semiotic artifacts

Deacon overall proposes for his autogen an autocatalytic network which will, in addition, also generate some by-products which will serve as boundary/capsid units. Both of these processes (self-assembly and reciprocal catalysis) co-create each other in the following manner. If there's some autocatalysis initially, in a similar form to



with A and D being some substrate, B and C being some catalysts which reduce the activation energy of the correspondent reactions, and χ some by-product which eventually polymerizes to form a capsid/boundary through self-assembly, these two processes "reduce eachother's underminings". Typically autocatalysis would lead to its own cessation, either through complete consumption of substrate or eventual diffusion of the catalysts away from eachother, such that the formation of the capsid limits such diffusion. Although, there would obviously be a problem with lack of substrate once such capsid is fully formed without allowing permeability of the substrate, and so Deacon relies on external conditions, namely the availability and affinity of such substrate to capsid units, which would eventually lead to the (partial) collapse of the capsid through distabilization (here taken mainly as competition between capsid-capsid linkage and capsid-substrate linkage, in order to minimize the energy of the system). This type of process would lead some of the contents being "spilled" out of the container, and allowing for formation of new

units if there's enough catalytic precursors to form the autocatalytic network again. It would therefore be a form of self-replication. The availability of substrate in the environment would be an indicator, in a proportional manner, of good conditions for replication, as units that would eventually break through substrate binding to the capsid, in these high conditions of substrate concentration would be very likely to assume again auto-catalysis.

Deacon obviously states this isn't enough, because there's a lack of a regulation which isn't directly affected by the system's dynamics, which are given by the system's constraints (almost like the representation of the system, within the system). In this manner, there should be an offloading of the system's constraints onto an almost immutable component or process. Deacon then states that nucleotide polymers which would first be taken for energetic purposes, with examples of GTP and ATP, as energetic intermediates which would be coupled through (de)phosphorylation to other reactions, would eventually serve in the inert state (that is the state where the container is fully formed and impermeable) as an immutable structure, such that individual nucleotides polymerize in a somewhat unbiased manner (regardless of the base in such nucleotide), but would eventually have a clear bias in the way catalysts would preferentially (according to shape complementarity between themselves and the nucleotide polymer) bind to some regions of the sequence of the polymer and not others. In this manner, it would control which catalysts are closer to each other and which aren't, and by that would be promoting/preventing reactions that are part of the autocatalytic network, essentially reflecting the systems constraints, or better put offloading the system's constraints onto these nucleotide polymers.

My main crux with it, is again on the necessity of external perturbations for the autonomy (or lack thereof) of the system. A disruptor of the boundary would presumably also be beneficial for such system, as it would allow selective and partial permeability, thus inducing some variability to the system.

Organisms are inherently different from other dynamical systems

I'm having severe difficulties putting everything into "clean plates" but nonetheless let's observe the fact, that even if one takes a more cybernetician approach into describing a certain system, essentially separating the dynamical system variables (which are amenable to dynamical laws) and control variables (which change the dynamical ones), a fundamental point is still being missed. There's still this separation of operator/operand. There's no such thing as pre-stating the evolution of a system which can be described as an organism. No matter how much one tries to employ similarities to systems which are chaotically determined (e.g. N-body gravitational system), these can still be presented, although not in a closed form. Organisms show a different type of causality,

a circular one. Perhaps, Weber and Varela put this into something clearly better than anything I saw before in "Life after Kant: Natural purposes and the autopoietic foundations of biological individuality", particularly in section 3.4. Not wanting to butcher something, that is clearly better than anything that I could put together on these themes, here are a few excerpts:

The key here is to realize that because there is an individuality that finds itself produced by itself it is ipso facto a locus of sensation and agency, a living impulse always already in relation with its world. There cannot be an individuality which is isolated and folded into itself. There can only be an individuality that copes, relates and couples with the surroundings, and inescapably provides its own world of sense. In other words by putting at the center the autonomy of even the minimal cellular organism we inescapably find an intrinsic teleology in two complementary modes. First, a basic purpose in the maintenance of its own identity, an affirmation of life. Second, directly emerging from the aspect of concern to affirm life, a sense-creation purpose whence meaning comes to its surrounding, introducing a difference between environment (the physical impacts it receives), and world (how that environment is evaluated from the point of view established by maintaining an identity).

The organic coupling and change must, according to its self-constitution, be always directed to maintain the process of self-realization. An autopoietic system is necessarily referred to itself: its actions consist in establishing the dynamical processes of staying alive. Stimuli from outside enter the sphere of relevance of such a unit only by their existential meaning for the keeping of the process of self-establishment. They acquire a valence which is dual at its basis: attraction or rejection, approach or escape. Form, then, is not just an abstract goal in a genetic program, but a material task to fulfil from moment to moment. The genetic program influences form, but only in being interpreted by the soma according to the actual needs of self-maintenance. Without the individuality of the living body the program is nothing – a fact that runs counter to the Dawkinian conception where bodies are machines acting teleonomically to unfold the underlying program and to maintain it (here the genome has the status of an idealistic principle of reason creating artifacts).

Conversely, if we follow the autopoiesis-Jonas inversion, if we accept autopoiesis as embodied teleology, we reintroduce the subject into biology. The separation of the realm of pure natural science from the realm of values, so popular since neokantianism (Rickert

1920), has to be abandoned; instead a theory of embodied meaning has to be reintroduced into the science of the living, paying central attention to categories as value and subjectivity. By defining itself and thereby creating the domains of self and world, the organism creates a perspective which changes the world from a neutral place to an Umwelt that always means something in relation to the organism. Organisms can be said to transcend the neutrality of pure physics and to create their concern. Only this organic perspective actually has the status of "world," only this is real, because the living can only act in the form of such an intentional world. Life is thus always subjective in the strong sense of the word.

The difference between environment and world is the surplus of signification which haunts the understanding of living and of cognition, and which is at the root of how the self becomes one. . . . There is no food significance in sucrose except when a bacterium swims upgradient and its metabolism uses the molecule in a way that allows its identity to continue. This surplus is obviously not indifferent to the regularities and texture (i.e., the "laws") that operate in the environment, that sucrose can create a gradient and traverse a cell membrane, and so on. On the contrary, the system's world is built on these regularities, which is what assures that it can maintain its coupling at all times. (Varela 1991, p. 86).

. . . for the first time within being the difference between substance and form, which is a pure abstraction when applied to the inorganic, becomes a real distinction. This implies a complete inversion of the ontological relationship: Form has become the essential, and substance has become the accidental. (Jonas 1973, p. 125).

Necessary then are the material compounds of an organism, their incessant input and their unhindered supply. But this necessity again is governed by a principle of autonomy – or, as Jonas says, freedom: the fact, that a living system is able to become an ontological center, that it is able to organize itself into a form that is not explainable by the features of the underlying matter (the pure necessity) alone. This autonomy then is nothing other than true teleological behavior. This autonomy has to do with the ever existing gap between the realization of the living and its underlying matter. Because form that desires itself in a purposeful manner is happening only in matter to which form is not its entropically "natural" state, there is always the possibility, and final certainty, of death. It is this existential situation that is emphasized by Jonas: the teleological, circular, self-referential movement of the living.

To live means to say yes to oneself emphatically as the basic movement of existence, because existence is always existence of form on and against pure matter. To speak of freedom or autonomy thus directly links the biological sphere with a teleological account of ontology. On a material, concrete level we can observe in the organism the flip side of mechanical causality, a final causality as the basic process of life itself – the establishment of an identity. But this happens not by revising physical laws for particle-interactions in special application to organisms, nor by imposing an extra-mechanical entelechy. It is rather the “subject-pole” that is the organism in its autonomy, which changes linear causality by structuring matter in the process of self-realization to maintain itself as this very process.

How does then, a system like this emerge in the first place?

Emergence of physics of symbols? (à la Pattee)

Perhaps a better statement for “How does then, a system like this emerge in the first place?” is: How does the symbolic (and respective semantics) association to the physical realization of the system begin? How does its *umwelt* begin?

We usually talk about all different types of (de)codifying processes in biology (particularly in molecular biology), e.g. transcription and translation, etc. However, this is a “clean story” of the problem on part of the interpreter system, on part of the organism. It is completely contextual, eventhough clearly conserved to a certain degree, given the inherent continuity that’s needed for an organism *to be*. However how does it start? Autocatalysis or reciprocal catalysis isn’t enough, and although Deacon’s autogen and similar approaches give us a window into such, isn’t this the *specific* case? What’s the *general* one? What’s the *general* case for the offloading of constraints onto semiotic artifacts, in a iteratively self-interpretative manner, as Deacon puts it?

On von Neumann’s universal constructor and its relation to autopoietic systems

This is mostly talking about Hofmeyr (2018) “Causation, Constructors and Codes”, and trying to get better intuition about a “bridge” between such universal constructor, Rosen’s relational approach relying on Aristotelian (be)causes (along with Hofmeyr’s development), and moreso the relation to autopoietic systems as described by Maturana and Varela, and the importance of what Pattee (and subsequently Deacon with his autogen) was talking about, about a system which self-constructs and which has the emergence of a symbol-matter relationship.

The universal constructor described by von Neumann, A , forms an automaton consisting of $A + \phi(X)$, under which $\phi(X)$ is a description to build the machine from component parts. For the necessity of the automaton to also pass on a description of itself (beyond merely itself, A), von Neumann also adds a description copier B and a controller C , such that the automaton becomes $(A + B + C) + \phi(X)$, with the new copy being $(A + B + C) + \phi(A + B + C)$. I'm not going to develop on the relations between such Aristotelian causes in the system, as that's already done by Hofmeyr, and I'm merely trying to relate this, to Deacon's autogen. The obvious point being, that the system needs to be completely self-referential. In that, the description of the system $\phi(X)$ to build should be reasonably immutable to the system's dynamics, and this is precisely what Deacon's autogen gets us. It is allowed by the "offloading of constraints onto semiotic artifacts", by Deacon's words, onto structures or processes which are largely immutable. There need to be such structures which code for the systems constraints (and therefore dynamics) and which can be decoded. Through the incorporation of nucleotides onto polymers, which initially have purely an energetic function, such sequence will inherently "bias" and constrain the system, through the specific affinities of the catalysts to each nucleotide in the sequence and therefore promoting (and preventing) certain reactions, merely by such catalyst proximity. Although, in the autogen the induction of self-replication is based on external perturbations, namely on the existence (proportionally) of substrate on the environment which will bind to the capsid (and compete for bonds), ultimately leading to capsid disruption, and the "spilling", if not atleast partially, of the contents into the environment. After this, through both coupled processes of reciprocal catalysis (akin to autocatalysis) and self-assembly, other units will be formed. Substrate concentration in the environment is a clear indicator of good (as a judgment value) conditions for self-replication. What I presume is missing, is the existence of internally generated perturbations which can aid the system from converging completely to the inert state, in the eventuality of there not being strong enough external perturbations. What are also needed are alternatives, as this to me would be one of the specific cases. Therefore one should look for the general case of how such self-constructing (or autopoietic) systems emerge, and how the corresponding symbol-matter relationship appears.

The ontogenesis of autopoietic systems

Now, the most fundamental problem is surely how such systems emerge. How does the first universal constructor emerge? When Rosen points out that the physiology (or how the system evolves/changes over time) is at least partially divorced from its ontogenesis, this surely also applies even if such system is treated in a relational manner. Won't ontogenesis tell more about physiology,

when compared to the inverse case? Wouldn't what we capture from such physiology tell us about the specific case of such ontology, and not necessarily about the general case? Won't some of the assumptions, that we take fundamental from the specific, be incorrect for the general (at least not fundamental)? This *is* the problem. And for the major part, this is seen, by us not being able to re-create anything even remotely close to what organisms are. Regardless of any so proclaimed advancement in the bio-sciences, arranging a system such that it becomes an organism (abiogenesis) is something that we are not even remotely close to. And mind you, this isn't the aim. The aim is to understand the ontology of the organism. The general case. Not the specific ones. Witnessing abiogenesis again would merely follow from it, such that it isn't the aim.

Deacon's autogen lack of internal perturbations

Still on this matter, having an autocatalytic network which produces itself, a boundary system (capsid units) and a disruptor system (by-products which compete for bonding with capsid units) using a certain substrate molecule, one inherently increases the sensitivity of the autogen to substrate concentration (good conditions for self-replication) and other relevant external perturbations, whilst "guaranteeing" some variability being added to the system, on top of the system not converging completely to the inert state (in the absence of strong enough external perturbations). Mind you, that this would still employ Deacon's autogen having the capsid molecules bind and "respond" (proportionately) to substrate concentration in the environment, eventually leading to capsid disruption.

If there is low concentration of substrate in the environment, and therefore not enough to lead to the collapse of the capsid, there's at least some type of semi-permeability to the capsid, which inherently allows some variability to be put into the system. It's important to note that this is ratio of capsid/disruptor molecules produced should be one which still allows self-assembly of the capsid to happen in a reasonable manner. In instances of no substrate in the environment no capsid nor disruptor molecules are going to be produced, but some reactions which fall "outside" of the autocatalytic network will still happen. This of course, taking this rigid approach to classes of molecules (serving for boundary, capsid, and autocatalytic systems), which would presumably be unlikely, that is, be much more fluid and have the same molecule (in different conformations) perform different functions.

Capsid-Disruptor relation

Still regarding the last topic, such disruption and capsid systems could essentially stem from the same molecule, although with different conforma-

tions. For self-assembly to still be reasonable, the formation of the disruptor molecule, or the conformation, by that matter, would need to be more unlikely, such that its ΔG is superior to the capsid's one. Once, such conformation is achieved it would compete for bonds between capsid molecules, thus leading to (partial) disruption of the boundary. Even better would be for such ratio between free energies of the corresponding conformations to change given the "closedness" of the autogen. If the conformation of the disruptor was promoted in such conditions, the inert state would never be achieved completely, and thus complete reliance on external perturbations wouldn't be needed.

On a few ramifications of Deacon's autogen

I'm not going to digress too much into what's already been stated, particularly what seems to be a lack of generation of internal perturbations, which could account for a minimal form of autonomy, for example with such perturbations being generated as a by-product of autocatalysis. This would also increase the sensitivity of the autogen to selection, as now the capsid system wouldn't be the only one under "stress-testing" (selection for optimal reproduction and repair rates). The autocatalytic plus the self-assembly systems would be encountering larger variability, through the occasional permeation of molecules into the autogen through the capsid's disruptions formed by the internally generated perturbations (bond-competition mostly). This can presumably be achieved with a single molecule, both with alternating conformations, one which would polymerize and allow capsid formation, and another one which would compete for bonds with such polymers, and thus deemed as a perturbation. The likelihood of such conversions could also presumably be a function of voltage gradients accounting for the material "closedness" of the system, such that there's a decrease of the free energy relative to the conformation of the disruptor, when compared to the capsid conformation, and an increase when the opposite is the case. As such, we have a stronger signal with respect to the environment, than that given by Deacon for optimal self-replication conditions (i.e. high substrate concentrations).

Regarding the offloading of constraints onto largely immutable structures or processes, this is of course done in a random manner, such that it's a matter of selection over each autogen. Every now and then, there are going to be autogens with nucleotide polymers which represent the constraints of the autocatalytic + self-assembly reaction network well enough, such that the autogen is viable. This offloading of constraints is the crucial key to open-ended evolvability. Moreover, this offloading of constraints needs to happen, if one is to not get an autogen (coupling of autocatalysis + self-assembly) which exhausts itself, through molecular combinatorial catastrophe, where the number of necessary co-dependent reactions for the repair of the autogen increases linearly, but the one for cross-reactions increases geometrically. There's got

to be a transfer of the constraints which lead to the repair of the autogen onto a structure which is largely immutable to such dynamics, otherwise there's a risk of this network of constraints not being able to support itself. If this offloading of constraints onto nucleotide polymers is enough, is another question. But it should be clear that a form of this, is completely necessary, if one is to get more complex systems, which don't derive it from complete rearrangement of material (without coupling of physical and symbolic domains, that is).

On that, an excerpt of H. Pattee on von Neumann's intuition:

Why does a material structure need symbols to communicate or transfer its structure to other matter? The first reason, mentioned by von Neumann, is that any universal constructor that could assemble its material parts would function more efficiently if it replicated from a symbolic description of its material parts rather than replication by material self-inspection of its parts. I have not found a strong physical or logical support for this efficiency argument, although it sounds plausible. In any case, as any evolutionist would point out, replication by symbolic description must be superior to material self-inspection because it survives. The second reason, indirectly suggested by von Neumann, is that new descriptions, being simpler, are more likely to arise than corresponding new material constructions. This is also not a general physical or logical argument, but again, it is plausible from our knowledge of molecular genetics and evolution.