An organisational systems-biology view of viruses explains why they are not alive


Published in:
Bio Systems

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
Copyright 2020 Elsevier B.V.
This manuscript is distributed under a Creative Commons Attribution-NonCommercial-NoDerivs License (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits distribution and reproduction for non-commercial purposes, provided the author and source are cited.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Open Access
This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: http://go.qub.ac.uk/oa-feedback
An organisational systems-biology view of viruses explains why they are not alive

Keith D. Farnsworth*1

1School of Biological Sciences, Queen’s University Belfast, 19 Chlorine Gardens, Belfast BT95DL, UK

last revised March 22 2022, including commentary from readers

Abstract

Whether or not viruses are alive remains unsettled. Discoveries of giant viruses with translational genes and large genomes have kept the debate active. Here, a fresh approach is introduced, based on the organisational definition of life from within systems biology. It views living as a circular process of self-organisation and self-construction which is ‘closed to efficient causation’. How information combines with force to fabricate and organise environmentally obtained materials, given an energy source, is here explained as a physical embodiment of informational constraint. Comparing a general virus replication cycle with Rosen’s (M, R)-system shows it to be linear, rather than closed. Some viruses contribute considerable organisational information, but so far none is known to supply all required, nor the material nor energy necessary to complete their replication cycle. As a result, no known virus replication cycle is closed to efficient causation: unlike cellular obligate parasites, viruses do not match the causal structure of an (M, R)-system. Analysis based in identifying a Markov blanket in causal structure proved inconclusive, but using Integrated Information Theory on a Boolean representation, it was possible to show that the causal structure of a virocell is not different from that of the host cell.

* k.farnsworth@qub.ac.uk
**Keywords:** \((M,R)\)-system, autonomy, circular causation, IIT, virus replication, autopoiesis

1 **Introduction**

The first half of 2020 has seen one particular virus (SARS-Cov2) dominate world news, so much that viruses appear to be at the forefront of public interest in biological research and in this context an old debate has reemerged: “Are viruses alive?” According to an informal survey (Racaniello 2014), expert opinion remains divided roughly a third each between yes, no and don’t know. This is not surprising given that the debate seems still to be resolved. Eleven years ago, an emphatic statement was made against including viruses among the living (Moreira and Lopez-Garcia 2009), quickly countered by (sometimes indignant) responses of matching boldness (Claverie and Ogata 2009; Hegde et al. 2009) and more nuanced responses (e.g. Forterre 2010b). The discovery of giant viruses (Raoult and Forterre 2008; Abergel et al. 2015; Claverie and Abergel 2018), especially the Pandoraviruses, having genome sizes reaching that of parasitic eukaryotes (Nadège et al. 2013) and Tupanviruses with their batteries of translational genes (Abrahão and et al. 2018; Rodrigues et al. 2020) has further stirred the debate (e.g. Claverie and Abergel 2010; Abergel et al. 2015; Brandes and Linial 2019). It also attracted philosophers of science who having analysed the debate, concluded that it is misguided (van Regenmortel 2016; Koonin and Starokadomskyy 2016). Whether or not viruses belong within the category of living has again become highly topical and contentious.

The answer, of course, has as much to do with how we define life as it does with the nature of viruses and that is the main criticism the philosophers had of the debate in virology. For (van Regenmortel 2016), the idea of viruses as a form of life is no more than a misconception (at best a vivid metaphor) brought about by the liberal
use of anthropomorphic expressions in virology. He quotes the well known virology
textbook (Flint et al., 2009), in which (for emphasis) the authors state that “viruses do
not actually do anything”: even the orthodox view acknowledges that they are passive
genetic parasites (citing Lwoff (1957) for this). With such confident statements, that
might have been the end of it, but it was not.

For a start, many virologists now consider the whole replication cycle of the virus,
insisting that the virus should not be confused with the virion and that to do so is
equivalent to exclusively focussing on the spore stage of bacteria, or (more obliquely)
on pollen. It does not help to say, as some do, that viruses are on the boundary of
life, first because that does not answer the question and second because it pre-supposes
a boundary between life and non-life, where none has yet been agreed. The “what
is life?” debate is arguably even more contentious than the question of viruses and
certainly older, so there is a danger of jumping out of the virology ‘frying pan’ into the
metaphysical ‘fire’ by addressing that head-on. It may be, however, that the question
of viruses has not found consensus precisely because the most fundamental and general
understanding of life has not yet been given due prominence in the discussion (e.g. the
survey of the topic by Herrero-Uribe (2011) has received little attention to-date). This
understanding is that life is the process of enacting closure to efficient causation (Rosen,
1991), meaning that a living system is the cause of itself. This is an idea initially
classified by Immanuel Kant (Ginsborg 2006 Gambarotto and Illetterati, 2014),
given rigorous definition by Robert Rosen, (1991) and practical interpretation as ‘every
catalyst necessary for life is produced by the living system itself’ (paraphrasing Cárdenas
et al. (2010), referring to Kauffman (1986)), placing it at the heart of systems biology
(Westerhoff and Hofmeyr, 2005) as a particular approach within it: the organisational
approach (Bich and Damiano 2012 Moreno and Mossio 2015).

This paper will proceed by first adding a physicalist analysis of cause to this, essen-
tially cybernetic explanation, showing how it discriminates life from non-life, then by
considering contemporary definitions of the virus to arrive at a test for whether these biological entities can be considered living in any known circumstances. The main point is that the study of viruses sheds new light on the nature of life itself.

2 What is life?

In posing this question, physicist Erwin Schrödinger inspired the deep scientific study of what it is that biologists examine, with the realisation that one had to reach beneath biological empiricism to find an answer. Despite that, biology textbooks commonly provide a list of attributes for living organisms: reproduction, metabolism, etc. (e.g. Soloman et al. 2002) and this is the standard approach in determining what is alive (Van Regenmortel 2010). It is far from satisfactory, since many things generally agreed to be non-living posses at least some of the attributes (fire, some computer algorithms etc. (Cleland and Chyba 2002)) and many organisms, not least viruses, lack some of them. (Rosslenbroich 2016) reviewed properties that have been proposed as indicative of life, but it remains the case that we cannot identify a boundary between living and non-living by ticking off the set of attributes, since it is unclear what subset of these is necessary and sufficient. In a well known objection, interspecific hybrid organisms such as mules would not qualify as living because they cannot reproduce and also then, do not evolve. This highlights the difference between identifying an individual organism as alive and considering a class of organisms as potential members of the living (Koonin and Starokadomsky 2016). An organism may be dead but be a member of a class that has the attributes of life and a thing may have the attributes, but not be alive, e.g. some autocatalytic chemical systems (Segrè et al. 2000; Zepik et al. 2001) and their hypothetical simulations (e.g. Hordijk and Steel 2004; Hordijk et al. 2012; Markovitch and Lancet 2014). The objection that some of these cannot evolve by natural selection (Vasas et al. 2012) is not decisive because evolvability is an attribute of all ensembles
of imperfectly reproducing entities which compete over a limiting resource, so it cannot identify a boundary between a non-living state of matter and a living system (Bruylants et al., 2010), on the contrary, it must span the transition between them (Nghe et al., 2015).

Even those with a more synthetic (as opposed to reductionist) frame of mind have set attribute requirements, such as enclosing membranes (Damiano and Luisi, 2010) and ribosomes. Requiring ribosomes for defining life is based on the three kingdoms of life proposed by Woese et al. (1990), though the RNA-first hypothesis of life’s origin allows for pre-ribosomal biology (Benner, 2010). Raoult and Forterre (2008) and Forterre (2010a) offer a counter-argument which includes viruses along with ribosomal organisms among the living. Cornish-Bowden and Cárdenas (2017) emphasised that the last common ancestor, LUCA, was not necessarily, or even likely to be close to the origin of life and to this extent LUCA tells us little about the transition from proto-life to life proper and, as they say there, “It hardly matters whether giant viruses are regarded as alive or not, because it is impossible to believe that life started with a self-organizing system with many proteins”: in other words, life cannot be defined by a threshold in molecular richness either. All of the ‘list definitions’ so far proposed are contestable (Piast, 2019; Bich, 2019) and mostly exclude viruses.

2.1 Life as organisation: the organisational biology approach

The organisational approach (Bich and Damiano, 2012), a strand within systems biology that is gathered under the heading of “current theories of life” in a substantial recent review of the topic by Cornish-Bowden and Cárdenas (2020), holds more promise as it defines life as a process enacted by a physical system: focussing on the difference between the active process of being alive and the passive (e.g. decay) process of being dead. The process of living counters the second law of thermodynamics by maintaining (and, as a by-product of success, reproducing) the integrity of the very system that enacts
the process. Rosslenbroich (2016), quoting Hofmeyr (2007) (p. 217) provides a good summary: “for systems biology, the defining difference between a living organism and any nonliving object should be that an organism is a system of material components that are organised in such a way that the system can autonomously and continuously fabricate itself, i.e. it can live longer than the lifetimes of all its individual components. Systems biology, therefore, goes beyond the properties of individual biomolecules, taking seriously their organisation into a living whole.” Self-referential systems are highly characteristic of life (Louie and Poli 2011). In the face of the second law of thermodynamics as well as a variable environment, self-maintenance implies both self-regulation (multiple homeostatic processes) and continuous (or at least frequent) re-construction of all of the systems parts: autopoiesis (Luigi 2003; Varela et al. 1974; Zeleny 1981) (these are not the same - see e.g. Bich et al. (2020)). This in turn requires the system to complete at least one thermodynamic work cycle (Kauffman 2000); i.e. it must export entropy to its environment by degrading energy (more precisely, transform free energy) to counter the second law in order to do work (in the thermodynamic sense). The work obtained from the closed thermodynamic cycle is realised as constrained (chemical) forces that together constitute the anabolic processes of self-maintenance - see discussion of work-constraint cycles in (Moreno and Mossio 2015, Section 1.2.1). By this, the organism assembles its body parts from material found in its environment (anabolism) and breaks down degraded parts (catabolism) to excrete them. These activities do not necessarily have to happen all the time, nor all at the same time, but they all have to happen at least some of the time during which the system can be claimed to be alive. The general concept of the organisational approach is summarised by Rosen’s (M,R)-system theory (Rosen 1985, 1991, 2000), in which processes are abstracted to categories, in the mathematical sense. Rosen’s ideas have been developed further by several authors, notably here, Louie who in 2013 Ch.13 applied it to “Relational Virology” and Hofmeyr, who has provided a concrete description of the cell as a hierarchical causal cycle Hofmeyr (2017): these
and related insights will be used further in the present work.

In the organisational approach, process is usually described in cybernetic terms, allowing it either to be a cybernetic model of the organisation of material transformations, or literally an algorithm (e.g. a computer program). Let us set the latter aside since it will next be argued that life can only exist as a material system. The cybernetics of life can in principle be embodied by any appropriate substance (as long as it works), but a computer algorithm (such as part of Conway’s Game of Life (Gardner 1970)) conceived and written by a human operator and running on a manufactured computer does not qualify because it has no natural independent existence: it is no less an artefact of human technology than a lightbulb.

2.2 Information embodiment and processing

All known life is a cybernetic process embodied in material: it is an integrated combination of relationships among diverse molecular components. (We will see why this must be so when both information and matter are identified as the ingredients of biological function). Embodied information underlies this diversity of molecular species and all the relationships among them. Embodied information is the pattern in space (and time) of ensembles of basic components (typically atoms), consistent with Landauer’s (1996) principle that all information is physical - see also [Karnani et al.] (2009). This is not merely conceptual: using information theory, [Jiang and Xu] (2010) have calculated the amount of information that is embodied in biological systems such as viruses and bacteria as a whole (taking a topical example, the bat coronavirus Rp3/2004, embodies 57720 bits of effective information in a genome of 59472 bits, coding 13 different proteins). Crucially, though, the information [Jiang and Xu] (2010) counted was only enough to reconstruct a virus given the amino acid and nucleotide building blocks: no virus contains the information needed to make these, they are given by the host cell. The information embodied by the shape of molecules can be estimated from their structural topology [Rashevsky]
and that of the nucleotides in RNA and DNA has been calculated by Sarkar et al. (1978), with other molecules and a more general treatment provided by Bonchev (1979, 2003); famously, Morowitz (1955) calculated the total embodied information of a typical bacterial cell to be $4.6 \times 10^{10}$ bits.

Embodied information is a familiar idea in relation to ‘information polymers’, but much more general: the type on printed pages, the magnetic stripes of hard disks, the charge variations in silicon memory chips and the electron cloud shapes of all molecules and other physical entities embody information in the spatial arrangement of their parts (Hazen, 2009; Rashevsky, 1955). Known life is information that is embodied in molecular shapes, in the act of processing information by pattern matching to synthesise, replicate, detect, disassemble and organise itself as a system composed of the material parts which embody the information it processes (Farnsworth et al., 2013). The ‘lock and key’ mechanism underlying much of biochemistry (not just receptors and ligands) exemplifies embodied pattern matching: steric and charge-distribution complementarity among molecules finds the maximum mutual information among molecules. The information embodying pattern of a physical entity is termed its form in what follows (see also Cademartiri et al. (2012) for discussion of the role of shape in self-assembly).

The information embodied as a particular configuration of molecules of a biological system at a particular time can be regarded as its global system state at that time. This is the combination of the form of its genome and the form of the set of all its other molecules combined. The number of possible states was termed the biological entropy by Jose (2020), who specified it as the product of number of possible genomes and the number of configurations of sensory states of the system that could embody time-dependent information about itself and its environment (where in this context ‘sensor’ means a set of molecules whose state depends on the the states of other molecules in the system). Jose (2020) summarised the total information capacity of a hypothetical population of organisms, with a genome encoded by an alphabet of $X$ base-pairs ($= 4$...
in known life) and length $L$, and given $S_i$ different sensors $s_1 \cdots S_i$ for each $e_i$ of a total number $B$ of entities (sets of molecules) to sense, in which the $j$th sensor detects $P_j$ attainable and detectable levels (i.e. values) of $e_i$. The upper bound of the population information capacity he calculated as:

$$C_{\text{tot}} = X^L \left( \sum_{i}^{B} e_i \sum_{j}^{S_i} s_j \sum_{k}^{P_j} p_k \right).$$  (1)

This information capacity (which counts every possible configuration of organisms with the specified complexity) acts as a dynamic working memory for the system that is considered to be processing information. $X^L$ counts all mathematically possible genome sequences, far more than biologically meaningful, but by specifying a particular genome from among all $X^L$, the information of the genome is maximised in the Shannon (information entropy) sense - as calculated by [Jiang and Xu (2010)]. Epigenetic switching enables state changes within the genome of all cellular organisms (Holliday, 2006), opening the way for information processing, but for an individual whose genome constitutes a static instruction set (i.e. it is not susceptible to changes in the system nor the environment), implementation of the instructions is as an automaton: it is part of a linear causal chain. In this static genome case all dynamic information processing must be found in the interdependence (sensing) of the non-genetic molecular configurations (we could say cytoplasmic system within cells). If that is absent as well, we are left with a static information statement, which is the characteristic of non-living entities, contrasting with the dynamic information processing characteristic of life. In other words, purely genomic information ($X^L$) is only effective at the evolutionary scale (the focus of [Jose's 2020 study]), or when it is combined with cytoplasmic molecular forms (e.g. when a virus accesses its host cell’s molecular machinery).

Information pattern matching (e.g. the sensory processing, referred to above) is part of life, only if it is functional in the sense that it is a necessary part of a causal relation.
with the effect of contributing to the process of living as a whole (Farnsworth et al., 2017b). So information processing is only effective if it is causative and only functional if the cause is a contribution to the organisationally higher level process of life (Walker and Davies, 2013, 2017; Farnsworth et al., 2017a) (note that Mossio et al. (2009) more strictly defined biological function as causal relations subject to closure). One implication is that life is a nested hierarchy of control structures in which obviously lower level interactions exercise casual power over higher. The idea that higher levels can exercise causal power over lower (and the same) levels of organisation is still controversial, but supported by several key authors (Auletta et al., 2008; Ellis, 2012; Jaeger and Calkins, 2012; Noble, 2012; Walker, 2014; Walker et al., 2016). It is less puzzling when we consider the physical basis of causation to find that embodied information is an elemental component, along with physical force, of all that appears to be cause, as explained next.

2.3 The physical meaning of causality: form and function

The philosophy literature includes a large, venerable and diverse cannon on causation. For scientists, interest begins with Aristotle who separated the notion often translated as ‘cause’ into four categories: material, efficient, formal, and ultimate (final cause). In his account, causation involves all four because they are the four natures (or aspects) of causation. Most modern philosophers seem to pay little attention to this as by far the majority of their current work concerns efficient cause only, which is usually taken to be the only true cause (many believe the other three were not really causes at all). Efficient cause is the dynamic action of transformation, moving or converting one thing to another and it coincides with a rough ‘common sense’ idea about causation. Although several prominent philosophers agree with Bertrand Russel, (1912-1913) whose highly influential paper concluded that cause was a figment of the imagination, most practicing scientists still need and use the idea: as Nancy Cartwright argues, science would be “crippled” by abandoning cause (Cartwright, 1979).
the question of what is life: many philosophers challenge the fundamental basis for
the question, whilst others (closer to the practice of science) have defended it as an
operational concept (Bich and Green, 2017).

2.3.1 Efficient cause, incorporating formal and material cause

Let us here adopt a physicalist view, which claims that in the physical (material) world,
what we observe as efficient causation is always the action of a physical force (usually,
but not necessarily on matter). More precisely, the physical mechanism behind cause is a
transfer of a conserved quantity (energy, momentum or something more exotic like charge
or spin) in a material system according to the transference theory of Salmon (1984) and
Dowe (2000) which posits that there must be a spatio-temporally continuous connection
between one thing X and another Y involving the transfer of energy, momentum (or
other conserved quantity) for X to cause Y (the connection is via a force field). Physical
forces all either cause movement or its prevention and all have an orientation (direction)
in space. The realised movement (or prevention of it) is the vector sum of all the physical
forces acting on a particle at one time. In the absence of constraints the vector sum of
forces acting on each member of an assembly of particles is random and accordingly has
no (ensemble) effect, other than pressure (Fig. 1 A).
Figure 1: The informational building blocks of final cause. A) random forces are B) constrained by form (in this case a crystalline lattice). C) more information rich form, as in these bio-molecules can result in e.g. ligand-receptor binding as the shapes and electrostatic fields match (mutual information maximising) and a network of these may act as the components of a detection-signalling pathway (D), which has function in the context of e.g. homeostasis for the whole cell, implying a final cause ((C) can be regarded as a magnified view of the messenger molecule attaching to the ion channel’s receptor site).

Constraints acting on forces reduce the range of directions in which forces can act among an assembly of particles. Forces can only be constrained by the relative position of the particles from which they emanate; indeed it is these positions that determine the directions in which forces act. As stated earlier, the positioning of the constituent parts of a system is embodied information which here is termed form. When particles are positioned in a form that is not random (i.e. the information necessary to describe
it is mathematically compressible), then the form has a coherent spatial structure: its spatial autocorrelation is non-zero and more generally the form has non-zero spatial mutual information (which is what is being termed ‘coherence’ here) (Fig. 1 B). This is the basis for effective information (Szostak, 2003). It is effective because it constrains forces in a way that gives them its coherence: specifically the directions of the forces are correlated by the mutual information of the form. The result is that forces, no longer random and merely producing pressure, act with coherence so that they are available to perform work and hence functions (e.g. the cylinder and piston of a steam engine is a form which constrains the kinetic force of steam molecules to act in a coherent direction producing a functional motion). This coherent action is nothing other than what Aristotle termed efficient cause: the action that brings about a transformation (or resists it). Hence efficient cause can be interpreted as the constraint of physical forces by form: force acting under formative constraint gives efficient cause. An important example of this basic unit of efficient cause in practice is the physical configuration of atoms in biologically relevant molecules that, as form, constrains intermolecular forces to act in coherent ways (coherent because there is non-zero mutual information) with effects such as binding and its consequences such as conformational changes (Fig. 1 C).

Traditional material cause, deriving from the composition of substances either acting or being acted upon by efficient cause can be seen in modern terms as a ‘micro-formal’ cause, since it is formal cause at the atomic scale. When high level (inter-molecular) form connects several material forms together, it can become an effective subsystem of biological metabolism, or perception and/or action, such as the ligand-gated channel system (Fig. 1 D). It is then clear that efficient cause is the product of material cause (micro-form) and information (I), which must be embodied as form in a structure that is not transformed by the process (e.g. a catalyst). More formally put as a mapping,
Rosen (1989) suggested

\[ f : A \times I \rightarrow B \]

\[ (a, i) \mapsto b = f(a, i) \]  

(2)

to explicitly incorporate information into efficient cause, where it plays the role of formal cause, with reference to his relational diagram for an \((M,R)\)-system (Fig. 2 a), in which \(A \rightarrow B\) is the set of material transformations from \(A\) to \(B\). Hofmeyr (2007) recognised that since \(I\) is a contribution to efficient cause along with \(f\), it should be associated with \(f\), not \(A\), and so rewrote this as (his Eq. 4):

\[ (f, i) : A \rightarrow B \]

\[ a \mapsto b = (f, i)(a), \]  

(3)

which recognises information as the formal cause that, together with efficient cause, generates the mapping (note, \((f, i)\) is an element of \(\{f\} \times I\), the combination denoting \(i\) informs \(f\), where \(i\) and \(f\) are members of \(I\) and \(\{f\}\) respectively). This reformulation of mappings was developed much further by Hofmeyr (2018), where formal and efficient cause are either combined into a single entity (informed efficient cause) by a “choice mapping” that selects a particular \(f_i\) from a set of possible mappings, or act together as separate entities \((f, i)\). Hofmeyr (2018) provides biochemical examples of both situations.

In many biological processes, for example translation of mRNA into polypeptide, an organic code (the genetic code), instantiated as adaptors (aminoacyl-tRNAs), mediates between formal cause (mRNA) and efficient cause (ribosomes). “Code” is here used in the sense of Barbieri (2015): a set of arbitrary rules establishing a mapping between two independent systems, which in biological systems has the effect of “translating an organic sign into its biological meaning” Barbieri (2015) quoted in Hofmeyr (2018).

This description was shown in Hofmeyr (2018) to be compatible with Von Neumann’s constructor theory of self-reproduction (Von Neumann and Burks, 1966), which represents reproduction as \((P + Q + R) + \phi(X)\) where \(P\) is a ‘fabricator’, \(\phi(X)\) is the
‘blueprint’ (information content) of machine $X$, $Q$ is a ‘blueprint copier’, $R$ a controller and for self-reproduction, $X$ will be $(P + Q + R)$. That is, there needs to be a fabri- cator and information about what to fabricate and both have to be duplicated for self reproduction. Living systems conform to this arrangement by embodying $I = \phi(X)$ in the form of $(P + Q + R)$. The necessity for information to be embodied and the real- isation that efficient cause is the combination of constraint by form on configurations of matter make it certain that living things are necessarily material objects embodying organisational information.
Figure 2: a) Rosen’s \((M,R)\)-System drawn as an autocatalytic network (taken from Cárdenas et al. 2010, Fig. 1.C., in turn from Goudsmit 2007). Solid arrows represent material causation (e.g. chemical transformations) and dashed arrows show efficient causation (e.g. catalysis). We can interpret material causation as the configuration of matter plus the matter itself and efficient causation as the information embodied in form plus the electrical (chemical) forces that this information constrains to enact the material transformations. An alternative biochemical representation of this was developed by Hofmeyr (2017), summarised in his Fig. 7, which is reproduced below as Fig. 3. b) Another interpretation of the \((M,R)\)-System which emphasises the cyclic character and separate, but connected efficient and material causes (hierarchical cycle) - redrawn from Louie and Poli (2011), is just Fig. (a), unravelled. c) This simplified sketch of Hofmeyr’s biochemical representation, based on his Fig. 9. in Hofmeyr (2007) shows how closely it matches the \((M,R)\)-System, though different in derivation. Again, solid arrows depict material transformations and dashed represent catalysis. Hofmeyr emphasised that protein folding and self-assembly of supramolecular structures are an essential part of living autopoiesis, often neglected in more abstract representations. Metabolic enzymes are efficient cause for constructing the biochemical building blocks of the cell, including of themselves and ribosomes. Ribosomes, tRNA, mRNA and associated proteins (the translation system) are efficient cause for transforming the building blocks into functional components, including themselves. DNA and transcription have deliberately been left out in this reproduction- they complicate the model without adding anything relevant to the current discussion.
2.3.2 Final cause - the taboo we cannot escape

That leaves only final cause, which pre-supposes a ‘purpose’ and that is necessarily subjective since purpose can only be in the view of the agent under study: purpose cannot be defined without reference to the agency to which it belongs. The implied subjectivity might be thought enough to rule it out of science, but in the case of organisms (uniquely) it is possible to say “objectivity is achieved through recognising this inherent subjectivity” (Bueno-Guerra, 2018), through the application of von Uexküll’s Umwelt concept, because organisms at the very least create the appearance of autonomous agency. This appearance is shown to be substantial, not an illusion, when organisms are revealed as systems of ‘self-entailment’ (Rosen, 1985; Kineman, 2011), meaning that they exist by virtue of closing their loop of efficient causation.

In all cases other than for organisms, explaining actions by referring to the ‘viewpoint’ of the system is unscientific anthropomorphism, but uniquely in the case of living organisms, explanations are at best incomplete without such reference. Biology requires a richer causal language solely because of the peculiar attribute of organisms apparently being causal agents (Bich and Damiano, 2012; Friston, 2013; Froese et al., 2007; Kauffman and Clayton, 2006; Varela, 1979; Vernon et al., 2015). This causal agency arises whenever a system embodies autonomous functional information, in particular a homeostatic set-point (Farnsworth, 2018, 2017) (since functional information is causal information where the effect is a contribution to the process performed at the organisational level of the system that embodies it (Farnsworth et al., 2017a)). The autonomy of the functional information depends on there being circularity of causation in the construction of the system in which it is embodied: without the circularity, the functional information would be causally linked to (an effect of) the system’s environment. Indeed, it is only with circular causation that internal can be distinguished from external and only with that distinction can information be autonomously embodied by a system (Bertschinger et al., 2008; Bich and Damiano, 2012; Froese et al., 2007; Kirchhoff et al.).
Agency is only superficially accommodated by the ‘machine metaphor’ (Marques and Brito, 2014) in which actions are mechanistically determined by complicated sequences of molecular interactions which occur within, and are part of, the organism (e.g. Hawkins, 1984; Capra and Laub, 2012). In that sense, agency is a proximal cause, though it rests on underlying physical processes. For those who reject the idea of organism agency (i.e. organisms as the initiating cause), evolution is evoked to explain the successful functioning of the perception-action ‘machine’: every part of a machine performs a particular role within it and is therefore functional with the implication that it must thereby have a purpose. Natural selection has evolved the parts whose functions are no more than the ‘appearance of suitability’, selected by competitive replication, so ‘ultimate explanation’ for action (behaviour) is evolution according to these critics (Fiore et al., 2015). This is a thin argument: as Rosen (1985) pointed out, evolution is a phenomenon of life, not the other way around. Even though organisms are evolved to perform fitness enhancing actions based on their perceptions, we cannot escape the point that it is the organisms performing these actions, not evolution, nor the underlying physics (Farnsworth, 2018). In short, living organisms are unique in having agency and they need to be closed to efficient causation to achieve it (Moreno and Mossio, 2015; Mossio et al., 2009, 2013).

2.4 Closure and its consequences for life

Metabolic closure (Letelier et al., 2006, 2011) is the closing of a chain of efficient causation that leads to the maintenance of a living system through biochemical processes. Recalling that efficient causation requires both information-based constraint of forces and material (the source of those forces), it therefore means closure of informational constraints (Montévil and Mossio, 2015) and the processing of material (hence the bio-chemistry). More practically, this means that all the catalysts necessary for the life of
a system (organism) are produced and/or maintained internally by the system (using raw materials from its environment). The catalysts are produced by the action of one another, through the construction of their forms by assembling molecular embodiments of information that has already been embodied within the system (Fig. 3).

This is initiated by pattern matching through genetic transcription and translation, but also includes purely biochemical chains of (spontaneous) anabolic reactions. We can...
see how DNA and RNA provide a template (pattern) which is matched in proteins that
in turn function in the fabrication of other necessary proteins - and also the compo-
nents from which they themselves are made. Some of these proteins are the material
forms needed to maintain and replicate the DNA and RNA templates. That is the ba-
sis of the closed loop. Hofmeyr (2017) emphasises the causal separation between the
fabrication of unfolded, unassembled biopolymers (covalent chemistry) and their forma-
tion into functional components by supramolecular chemical processes, this enabled by
the highly specific chemical environment, including chaperone molecules, proteasomes,
splicesomes as well as small molecules; collectively the intracellular milieu. The milieu is
itself maintained by molecular transporters (transmembrane selective channels), which
themselves are assembled and made functional by the same processes. In the catalytic
transformation of nutrients into biopolymers we see $a \rightarrow b = (f, i)(a)$, where $i$ selects the
functional catalyst necessary for each and is materially embodied in the molecules of the
intracellular milieu. Life is necessarily physical and material, as well as informational.
Identifying life should therefore include requirements for the selection of material build-
ing blocks from the environment (nutrients), their processing into functional proteins
and the organisation and regulation of these processes into a closed causal loop which
results in reproduction (the copying of the material form, together with the organisational information, including its information template - the nucleic acid ‘blueprint’). Let
us now see to what extent known viruses match such a description.

3 What is a virus?

Our knowledge of viruses has progressed tremendously in the past twenty years, leading
many to consider revision of what we mean by the term ‘virus’. In reply to Raoult
and Forterre (2008), Wolkowicz and Schaechter (2008) claimed that the defining char-
acteristic of a virus is that it undergoes disintegration and reconstruction as entirely
separate stages of its replication cycle. Still, the standard definition provided by [Raoult and Forterre (2008)] stands: it is “a capsid-encoding organism that is composed of proteins and nucleic acids, self-assembles in a nucleocapsid and uses a ribosome-encoding organism for the completion of its life cycle”, even though that excludes viroids and endogenised genetic material (inserted into host genomes). [Claverie and Ogata (2009)] emphasised the diversity of what they considered organisms having a range of replication cycles, deeply rooted in the ‘tree of life’ - specifically not the virions for which the term ‘virus’ was first created and not mere ‘gene robbers’: whatever we call them, many have uniquely virus genes. The many giant viruses now discovered are remarkable in creating an elaborate “viral factory that resembles a eukaryotic nucleus” ([Suzan-Monti et al. 2007](#)) with which they deploy an impressive range of functional proteins ([Brandes and Linial 2019](#)). Of key interest among these are tRNAs, ribosomal proteins and other translation and transcription proteins, all coded within the virus genome ([Schulz et al. 2017](#)). None have been found with the full set required for independent reproduction, but the argument that viruses are incapable of reproduction without the host’s translational machinery has taken a few steps of retreat. Some giant viruses have been found with enough of their own transcription proteins to perhaps transcribe independently within the virus factory and also have some metabolic pathway genes (see e.g. [Schulz et al. 2017](#)), leading several virologists to say that they are equipped with “most functions traditionally attributed to cellular organisms, including: Protein translation, RNA maturation, DNA maintenance, proteostasis and metabolism” ([Brandes and Linial 2019](#)). For those virologists viewing viruses as ‘life’, they are united by having capsids but no ribosomes, while other domains of life have ribosomes, but no capsids ([Raoult and Forterre 2008](#)). This seems to imply an equivalence (hence substitutability) between capsid and ribosome, presumably unintended. According to the definition of life based on the theory of autopoiesis ([Luisi 2003](#) Varela et al. 1974), both an enclosing physical boundary and a self-creating synthesis system
are needed - not one or the other. Within the organisational approach, the “relational virology” of Louie (2013) precisely interprets the virus as an ‘entailment network’ (the interconnection of causal necessities) coupled to the entailment network of a host cell, strictly via genetic interaction - the replacement of genetic information in the cell. [Important conclusions: the virus contributes no material cause and its entailment network is not cyclic].

Viruses should not be considered exclusively parasitic as some provide considerable advantages for their hosts, in particular those phages that equip their prokaryotic hosts with defences against their eukaryotic host, increasing the virulence of the prokaryote. For example the phage Sp4 gives a superoxide dismutase to E. coli helping them survive oxidative stress, whilst phage lambda gives both an adhesin to promote adhesion to buccal epithelial cells and a new outer membrane protein that confers resistance to serum complement killing (many different host virulence enhancements are reviewed in Boyd and Brüssow (2002)). This leads some to think of the virus-host system as a composite holobiont, but if it were truly integrated as a whole, then we would more reasonably consider the virus not as a life form in its own right, but rather as a part of the chimera which includes an extra-cellular phase. In the extreme, the virus is incorporated as part of the host genome, entirely loosing its extracellular existence (e.g. as a transposon). The idea that viruses could be life because they have to be considered in combination with their host does not seem to be a logical defence in any of these cases because the virus loses its independent identity: it becomes a part of the host as much as any other genetic element (more generally Lopez-Garcia (2012) called this argument “alien to logic”). The concept of partial autonomy in genome replication, used in this context by Koonin and Starokadomskyy (2016), certainly accounts for the distinct identity of the replicating unit, but this is no more than a local peak or plateau in the mutual information of the genome of the host. The incorporation of viral genes into a host genome is most evident and advanced among those transposons having a viral origin, for
which the term ‘autonomy’ has the narrow technical meaning of possessing a transposase gene.

Many viruses with eukaryotic hosts will cause intracellular compartments to be made, within which viral replication and assembly takes place, shielded from host defences. A broad range of compartment types, from relatively indistinct viroplasm formations to the most organised viral factories have been identified (reviewed by den Boon et al. (2010) and Novoa et al. (2005)). In the few cases of giant viruses so far know, the viral factory can be a place where translational molecules of viral origin are highly expressed (Rodrigues et al. 2020), but so far, perhaps crucially, no viral ribosomes or functionally equivalent components have ever been detected (in their closing paragraph, Rodrigues et al. (2020) speculated that it was just a matter of time before they are). Also, the reproductive activities taking place within the viral factory require an energy supply and this is not provided by the virus: several with eukaryotic hosts have been observed to recruit host-cell mitochondria to the site (Novoa et al. 2005) or manipulate host metabolism to obtain energy (Chuang et al. 2017; Nagy and Lin 2020), as they also manipulate host metabolism to produce e.g. viral lipids (Rosenwasser et al. 2016). This leaves us where we started: a virus is a biomolecular system having many of the basic components of an organism, but lacking its own ribosomal machinery or any equivalent, it depends on a ribosome encoding organism to complete its replication cycle, (Raoult and Forterre 2008) as well as needing its host to supply energy and precursor molecules for reproduction.
4 Do any virus-like systems achieve closure to efficient causation?

4.1 Evidence in the virus replication cycle

To attempt an answer, the first thing we must do is interpret the replication cycle of the virus as a causal network. The general replication cycle of a virus consists of attachment, penetration, replication, assembly and release phases. For both attachment and entry, recognition of the host molecules is achieved by molecular pattern matching: when mutual information reaches a chemically determined threshold, the penetration stage is triggered. There are several kinds: entry may result from a conformational change in the capsid (in pore-mediated penetration); receptor mediated endocytosis, e.g. clathrin mediated, which recruits adaptor proteins from the host to help form a vesicle that carries the virus into the host cell; or the virus membrane may fuse with the host cell membrane (as in coronavirus). This stage may also involve signalling, but is generally thermodynamically spontaneous, even in the more complicated case of e.g. the T4 phage with its quite elaborate mechano-chemical system (having the appearance of a cleverly designed mechanism). Thus the first two stages are brought about by mutual information between the form of the virus and that of its host, presumably created by the evolution of the virus (perhaps co-evolution with the host). The virus DNA or RNA is then released into the host cytoplasm (via spontaneous chemical mechanisms that also differ among virus types). mRNA is needed for replication and in the case of positive strand RNA viruses (Baltimore class IV), this is directly available from the virus. By the current definition (see above), no virus has, or can autonomously create, ribosomes. Hence the viral mRNA relies on host ribosomes for translation. Picornaviridae (Class IV) are among those using an internal ribosome entry site (IRES) to enable host ribosomes to translate their RNA into a giant polypeptide, which in the first clear case of circularity, self-cleaves by internal proteases into functional proteins, one of which is the
RNA-dependent RNA polymerase. Another product of the polypeptide auto-cleavage is itself a protease which goes on to create the other functional proteins - a protease which acts upon itself. In terms of causal links, this amounts to viral formative information acting upon itself and being acted upon by part of the host’s formal information (from the ribosome). This causal arrangement is true for all known viruses, reverse transcription and the contribution of translation machinery by giant viruses included.

Speculatively, translation might be achieved by some (giant) viruses using entirely viral tRNAs, chaperones and associated enzymes (Abrahão and et al., 2018), but a source of ATP is required and in all known cases supplied by the host (Raoult and Forterre, 2008; Nagy and Lin, 2020). Finally, virion release is achieved through one of lysis, exocytosis or budding. In each case, material is recruited from the host to perform the release. Lysis usually involves the late translation of lytic genes using host material to construct the lytic agents; budding modifies and commandeers the host cell membrane and exocytosis (a normal cell process) is hijacked by some viruses (e.g. the α-herpesvirus pseudorabies), using cellular material and information.
In attachment / penetration, virus (V) and host (H) forms combine as mutual information (MI) leading directly to viral genes entering the host. Viral genes constitute information which acts as formal cause in conjunction with the host ribosome (the efficient cause) to transform materials supplied, along with the necessary energy, by the host (material cause), leading to the replication and assembly of new virions (template replication being repeated formation of MI). Viral genes, as formal cause, act on materials supplied by the host to either make lysis molecules that transform the host into a lysed cell, or form the structures needed for budding or exocytosis (which is a host function).

What we see in this generalised virus replication cycle, is that each stage is a mechanistic link of a linear causal chain that depends on both the virus and its host (fig. 4). In particular, the virus contributes functional information (embodied by its genome), but lacks both the necessary material and energy (for entropy reduction) to complete the physical replication cycle. The virus therefore influences efficient causation at each stage of its replication cycle, but without the material and energy supply it is not an
independent source of causation at any stage. This lack of independence in generating
causes precludes it from achieving closure to efficient causation for the simple reason
that it is not a sufficient source of cause. The virus, taken alone, lacks both Kauffman’s
thermodynamic criterion ([Kauffman, 2000] and Rosen’s ‘closure to efficient causation’
criterion ([Rosen, 1991]) for defining life. In partnership with its host organism, the virus-
host complex meets these criteria, but [Lopez-Garcia (2012)] was surely right to call that
notion illogical and invalid when considering the living status, specifically, of the virus.

The virus cannot control the environment needed for reproduction (it relies on the
homeostasis of the host cell), nor can it select the necessary materials from its environ-
ment (it relies on the host cell to provide these). The fact that it lacks the genes for
ribosomes is not of critical importance, even though it is part of the current definition
of virus. That is because even with ribosome coding, it would remain an information
parasite since none of its information would be effective (causative) without appropriate
material to constrain. For the same reason we do not accept as living any so-called au-
tonomous robot which depends on another system (people) to make its constituent parts
(this being true even if the robot were one that assembles its own parts, since it would
rely on people to extract and process the raw materials - a point made by [Hofmeyr’s
(2007) factory analogy]).

4.2 Ribosomes and origin hypotheses: lack of closure is an efficient
aparitic strategy

The lack of any coding for ribosomes raises an interesting question, because in principle
there is no impediment to the required genes being acquired and incorporated. Depending
on which of several hypotheses about the origin of viruses is true, ribosome genes
may have been jettisoned (according to the ‘regression hypothesis’), or never present,
following either the ‘escaped genes’ hypothesis or the ‘early virus’ hypothesis in which
viruses may have preceded cellular life in the evolution of early replicators - see [Farias
}
et al., 2014}. Given this, it is possible to speculate that proto-life (e.g. the RNA world) took two different courses: one developing via primitive ribosomes into cellular life and the other, lacking any translational machinery of its own, rapidly developing alongside as an RNA-based information parasite. In this scenario, leaving translation to the host may be the virus solution to Eigen’s paradox: no efficient enzymes are possible without accurate information templates, but no accurate information templates are possible without efficient enzymes (described with historical detail in Cornish-Bowden and Cárdenas 2020). It is now understood that ribosomes evolved by a series of additions to the translational core containing the peptidyl transferase centre (Fox 2010; Petrov et al., 2014), which is considered to be the oldest translational system (Petrov et al., 2015), hence the bridge between a proto-biotic RNA world and the biotic ribonucleoprotein world (Farias et al., 2017), thus preceding genetic sequence-based template reproduction (Farias et al., 2014). Coding the ribosome has perhaps never been part of the virus strategy because it is more efficient to rely on the host to go to the expense of maintaining such error intolerant and relatively large structures (requiring more than the error catastrophe limit of circa 200 base-pairs (Maynard Smith and Szathmáry 1995, pp 44-49)). Even if primordial replicators were the origin of viruses, what is left of them in modern viruses does not possess closed causation, since a host organism is always necessary to complete the replication cycle. Under the other two popular virus-origin hypotheses: if viruses are stripped down former organisms, the same holds and if they are escaped genetic replicators, again they have never been closed to efficient causation.

### 4.3 Can closure to efficient causation be quantitatively detected?

The only way to detect and perhaps quantify cause is through intervention (Pearl 2009; Woodward 2003, 2016). So far the methods offered have almost exclusively concentrated on linear chains of causality, or systems that can be represented by directed acyclic graphs.
4.3.1 Markov blankets

Friston (2013) proposed a Bayesian statistical approach (used for time-series data) to identify the characteristic organisational structure of life with a Markov blanket. It relies on partitioning causal subsystems which describe causal graphs of system states (hence not easily extended to ontological causal problems). The Markov blanket approach was initially proposed in the context of Bayesian networks of statistical relationships by Pearl (1988) and has been applied to the study of self-organisation in neural networks, (e.g. Kirchhoff et al., 2018). Specifically, a Markov blanket is a set of vertices in a directed probabilistic graph, which separates two other sets by conditional independence (one set is independent of the other, given the blanket). It can therefore be used to imply the existence of internal states, distinct from external states, such that internal states are not causally dependent on external ones. This is a tempting prospect because the causal boundary identified by a Markov blanket could coincide with the necessary internalisation of causality of autopoiesis and autonomy and entailed in cyclic causality (Palacios et al., 2020). Unfortunately, Bayesian networks are meaningful only for directed acyclic graphs, so although Friston (2013) used them to show how a Markov blanket emerges from a control system that seeks to minimise free energy by active (and embodied) inference (Conant and Ashby, 1970), his paper did not show that the Markov blanket indicates closure to efficient causation; indeed his analysis referred to the self-regulation of a system connected to a variable environment, not the self-making of that system. Despite that, we can usefully interpret the cell-virus system via a cyclic graph model (Fig. 5).
In principle, we could factorise the probability \( p \) of the graph, taking the directed edges as conditionals, e.g. for \( B \):

\[
p(B) = p(P|\{B \land A\}) \cdot p(R|N),
\]  

(4)
but the cyclicity makes the full factorisation of \{N, A, R, B, P, E\} a tautology in which \(p = 1 \forall \) nodes; i.e. cyclic Bayesian networks do not make sense. We can, instead, treat the network as a Markov random field (MRF) by abandoning the directedness so that edges of the graph represent potential functions (hence for the following analysis we should ignore the arrows in the graph). Labelling the set \{N, A, R, B, P, E\} = \(S\) (and for clarity relabelling the component parts \(N\) and \(A\) collectively as \(C\) and the nutrients as \(n\));

\[
p(S) \propto \phi(C, n) \phi(R, C) \phi(B, R) \phi(B, P) \phi(P, B) \phi(P, C) \phi(E, P) \phi(C, E),
\]

in which each \(\phi\) is a potential function relating variables in the factorisation. In general, factorisation of an MRF (with \(\alpha\) as a constant) has the form:

\[
p(S) = \alpha \prod_{i \in Q} \phi_i(x_i),
\]

where \(Q\) is the set of cliques, defined as a subset of all the nodes in the graph \((G)\) for which every distinct node is adjacent, i.e. for every pair of nodes \(u\) and \(v\) in the clique \(Q\), \(u \neq v\) and the edge \(uv \in E(G)\), the edge set of \(G\), so all the nodes in \(Q\) must be connected by an edge in \(G\). Identifying the cliques in the cell-cycle graph, reduces Eq. 5 to:

\[
p(S) = \alpha \phi(C, n) \phi(R, C) \phi(B, R) \phi(B, P) \phi(C, P, E),
\]

and to include the virus, we just add its contributions:

\[
p(S_V) = \alpha \phi(C, n) \phi(R, C) \phi(B, R) \phi(R, V_R) \phi(B, P) \phi(C, P, E) \phi(E, V_E).
\]
With this, we can identify the *separating subsets* between all pairs of subsets in $G$ that make these subsets conditionally independent (conditional in the sense that only by specifying the values of the separating subset do we make one member of the pair independent of the other). In general, though, we can use the following local Markov properties of the MRF: i) all non-adjacent variables are conditionally independent given all other variables and ii) every variable is conditionally independent of all non-neighbour variables given its neighbours, which in turn defines a Markov blanket for every variable.

In other words, in the MRF, for any node, there is a Markov blanket consisting of all the neighbours of that node (i.e. all the nodes it is directly connected to). That is of considerable use in the design of artificial neural networks or the study of real neural networks when the values represented by nodes are measurable variables (as in Friston, 2013; Kirchhoff et al., 2018; Palacios et al., 2020), but in the present application, we are just borrowing the mathematical structure to identify dependencies in the cell-virus system. All we need to know about the local Markov properties is that they tell us that the presence and/or functioning of a focal node is entirely determined by specifying the state (presence or absence / functional or not) of its neighbouring nodes. Taking B, the ribosomes, for example, we can see that they are not functional if either or both of P and R are not functional, irrespective of whether vRNA is functional - and we do not need to enquire further into the presence of nutrients or functioning of metabolic enzymes.

More significantly, we can see that no function of the system is dependent on any of the viral contributions, other than viral replication, which in turn is strictly dependent upon them, e.g. viral reproduction strictly depends on the production of N (nucleic acids). This is of course just a formal way of saying that the virus is strictly dependent on the cellular host, but the host is strictly independent of the virus: we have not advanced much by using an MRF model.

To be fair to those pursuing the Markov blanket approach, the acyclic restriction can be lifted by explicit use of a dynamic system model (clearly $\dot{x} = f(x)$ is causally cyclic.
but solvable). For example, Dynamic Causal Modelling (DCM) (Friston et al., 2003) enables dynamic causal analysis of Bayesian networks. Equations of motion have to be specified and the dynamic system allowed to follow its trajectory in time to reach an attractor, which then describes the causal relations throughout the dynamic network (directed cause-effect dependencies and conditional independencies) as a hypothesis which is tested against time-series data collected from nodes in the system. DCM therefore involves comparing rival plausible models of causal structure with observed time-series of variables from within the system (an approach demanding tremendous detailed specificity). Friston (2013) used a more general (stochastic) dynamic causal model, given a Markov blanket, to show the emergence of perception from an embodied control system operating by free-energy minimisation in the context of a varying environment that separates out as a set of external states $\Psi$, leaving internal states $\lambda \in \Lambda$ isolated by the Markov blanket that itself is partitioned into sensory states ($s \in S$) and active states ($a \in A$). The internal states self-organise in conjunction with the active states (following the free-energy minimisation of the sensory states) to become an embodied perception of external states (Friston, 2013, Fig.1). To apply such a model to a cell-cycle seems a daunting task and no result is presently available, but it can be noted that a virus is not obviously self-controlling, or seeking to minimise free-energy or any other potential function, nor is it obviously in possession of a Markov blanket.

4.3.2 Integrated Information Theory

Rather more promising for the present purpose is the analysis of causal graphs using Integrated Information Theory (IIT) (Tononi, 2004, 2008; Oizumi et al., 2014; Marshall et al., 2018; Hoel et al., 2016) (originally intended for understanding consciousness), because it has already proved practical in the quantification of causal independence in cyclic causal architectures and the identification of the internalised information associated with them (Albantakis et al., 2014; Albantakis and Tononi, 2015; Marshall et al.)
2018, 2017; Juel et al., 2019). Using this, the hope is that the qualitative question of whether viruses can be considered alive could be reframed as the quantitative question of how much of the virus replication cycle is causally independent of its host-environment - and how much it is a source of cause (as constraining information) in that environment. IIT determines the causal structure of a system by simulating its perturbation in every possible way (so is very computationally expensive). Its overall measure of integrated information (\( \Phi \)) gives the intrinsically irreducible causal power of the system as a whole, in the sense that if any partition of the system into two parts makes no difference to its cause-effect structure, the whole is reducible to those parts (hence the term ‘integrated’). One obvious test here is to partition a virocell (the intracellular form of the virus including its reproductive components - [Forterre, 2013]) into its virus and host cell parts to determine the causal integration of the whole.

The network of Fig. 5 (without virus) was translated into a discrete Markov Boolean system (Fig. 6) in which 1 (ON) represents ‘exists’ and 0 (OFF) represents ‘does not exist’. Nodes were all represented as AND mechanisms (using the language of IIT from Oizumi et al. (2014)), since the existence of each depends on all its inputs being from existing (ON) nodes. We should take care to remember that network models of this kind are designed to represent state dependencies among existing entities, rather than their existence or otherwise.
Figure 6: A Boolean network model of the system (Fig 5) for IIT calculation, using logical AND as mechanisms for all dynamic nodes, representing the requirement for all their inputs to be ON for them to exist. Note nucleic acids and amino acids are lumped together as components C. Nutrients (F for food here) and supplied RNA are considered external (provisions) so fixed ON (indicated by shading). The unshaded nodes represent internal mechanisms of the cell. Viral RNA (vRNA) and enzymes (vE) are external and identified by dotted causal links. vRNA and vE are fixed ON to represent a virocell (virus infected cell), otherwise the are fixed OFF.

Boolean networks follow inexorable dynamics from any initial condition (a starting state specified by the set of node values – e.g. for nodes R, C, E, P, B, we could start with \{1 0 1 0 0\}) and converge onto at least one attractor: either a fixed point where no further changes to states occur, or complex, where dynamics follow cyclic or chaotic variations ([Kaufman, 1969]). The dynamics depends on the update algorithm; in the simplest case this is synchronous (all nodes updated concurrently). Asynchronous models are usually preferred for biological network representations because typically each node has its own characteristic timescale, but in the present application which is rich in auto-reflexive relationships (causal looping), synchronous seems reasonable (we will soon see why it is
Various methods have been developed to reduce Boolean networks to their effective equivalent (Matache and Matache, 2016) by eliminating simple mediator nodes (single input, single output, e.g. node $E$ (without virus)) and ‘stablized nodes’ which reach a fixed point irrespective of timing and initial condition). Of most relevance for the present (quite simple) networks are the (widely used) algorithms proposed by Saadatpour et al. (2013) for eliminating mediator nodes and stabilised nodes. The logic of the network (Fig. 6) can be written as:

$$G = \{ R \leftarrow (C \land RNA); \ C \leftarrow (E \land F); \ E \leftarrow P \leftarrow (C \land B); \ B \leftarrow (R \land P) \}, \quad (9)$$

where $\leftarrow$ denotes one sided logical equivalence (e.g. $R \leftarrow C$ means $R$ copies $C$). If nutrients and RNA are given, then RNA and $F$ are fixed ON, so they do not affect the state of any AND gates, so can be eliminated. Further, we can see that $E$ is indeed a simple mediator and with RNA and $F$ eliminated, $R$ appears to be a simple mediator also, but because it depends on $C$, which in turn depends on $P$ (via the eliminated $E$) and also determines $P$, the network is only reduced to:

$$G = \{ R \leftarrow C \leftarrow P \leftarrow (C \land (R \land P)) \}, \quad (10)$$

from which the auto-recursion becomes clear as we see $P$ depends on $P$, $C$ depends on $C$ and $R$ depends on $R$ in a single nest of loops containing loops (hence the reduced graph has only one element). A Boolean transition table is easily made for the three nodes $P$, $C$, $R$, relating the states that follow every possible current state (from $\{0,0,0\}$ to $\{1,1,1\}$) and it shows that $\{1,1,1\}$ is a fixed attractor and all other states lead to the only other (also fixed) attractor $\{0,0,0\}$. Since the network dynamics have only fixed attractors, it is unaffected by the update algorithm timing, hence (as promised) the synchronous update algorithm is appropriate (see Appendix for details).
The IIT calculator (Mayner et al., 2018) was first given the complete network shown in Fig. 6, with the initial state of nutrients (F) permanently fixed ON; RNA fixed ON and all other (dynamic) nodes OFF. For the subsystem containing all but the nutrients node (which is external), the overall IIT was non-zero (Φ = 0.028), indicating the presence of intrinsic integrated information. Taking the subsystem of all internal nodes {R B P C E}, (excluding the RNA, assuming this to be an external, given from inheritance), with all but RNA initially OFF produced the considerably larger Φ = 0.125 (more detail is in the Appendix). Taking this subsystem with all nodes ON gave Φ = 0.3125. This case represents the living cell alone. Crucially for the virus question, adding vRNA and vE to the system in either case did not change the values of Φ or any of the concept ϕ values.

Using the (Saadatpour et al., 2013) reduction of the Boolean representation (Fig. 6) eliminated B and E to leave the closed looped system of Eq. 10 (see Appendix for details) which has only the ontological fixed attractors: {1,1,1} and {0,0,0} (either everything exists or everything does not exist), as does the complete network, of course. It is immediately clear that every part of the internal system is able to both affect and be affected by every other part, since no part or partition of the system acts the same way if any other part is separated from it, hence the system is an integrated intrinsically irreducible whole (Φ > 0). Significantly, adding the virus (vRNA and vE) made no difference to any of the Φ or causal structure results. Considering their role within the network, where vRNA is associated with node RNA and vE with node E, which could both be eliminated using the rules of Saadatpour et al. (2013), this should not be surprising. The IIT result quantitatively confirms that the virocell has no more integration of causal information than the host cell, i.e. the virus itself contributes nothing to the existence of the system, according to the model used here, though of course a virocell cannot exist without a virus: it either exists if there is both virus and functioning cell, or it does not, which is a very simple causal structure. That leaves all
the closed loop causality that makes life such a special phenomenon, firmly a property of the host cell.

These results seem quite conclusive, but it should be recalled that IIT was not intended to be used for ontological (existence) causal questions like this. What can be concluded is that in the causal network models presented, the role of virus contributions has always been ancillary to those of the host cell. These ancillary contributions make a combined biological entity - the virocell - by adding sometimes very considerable amounts of information to the system. But although measures of total information content suggest that the largest virus genomes rival those of the simplest cellular organisms, we know that total information is not particularly informative - it is what the information does that counts. A good measure of this is the total effective information contributed by the virus relative to that of the host: effective information being that which by constraining forces, generates cause. From the IIT analysis and the preliminary logic analysis and network reduction, it is quite clear that the Boolean representation of the host cell is rich in cyclic causality and that the virus contributes nothing to that, other than existence / non existence of a virocell, depending on the presence of a virus.

5 Conclusion

It is now clear that viruses are a very varied group of systems, some with information richness that could rival simple cellular organisms (which lack many genes thought necessary for prokaryotic life [Claverie and Ogata 2009]) and all deploy nucleic acid templates that can evolve, especially in response to the changing environment presented by their hosts. They contribute, sometimes considerable amounts of, functional information for the completion of their reproductive cycle at every stage, but never all that is needed other than for attachment and insertion stages. In particular they do not contribute sufficient functional information to support closure to efficient causation. Specifically, they
lack the ability to independently organise the creation of the necessary set of catalytic proteins (enabling formal cause to inform efficient cause) and to create and maintain the necessary local environment - the intracellular milieu that enables viral efficient causes to become functional through folding and self assembly. As a result they cannot achieve closure to efficient causation without considerable support from their host organisms (Fig. 7.A). In the abstract terms of a relational diagram, this was anticipated by Louie (2013), though it might be concluded from Fig. 7.B that Louie’s relational diagram of viral infection is insufficiently concrete in molecular terms to convince most virologists. Finally, we can see where the virus infection interacts with the cell at the more explicit level provided by Hofmeyr (2017) (Fig. 3 above) by comparing Fig. 3 with Fig. 7.A. Consistent with Louie (2013 Section 13.2), the link is found in the replacement of mRNA with an ‘impostor’ (Louie actually calls it a rebel) which becomes formal cause for the replication of viral polypeptides, though many viruses begin with ‘impostor’ DNA and even include their own transcription enzymes among the viral polypeptides, all of which can be accommodated by Fig. 7 in Hofmeyr (2017).
Figure 7: A. This is the same as Fig 2, but now with a virus incorporating itself into the system (shown in grey shading), for which the efficient cause is viral attachment and penetration (stage 1 of Fig 4, termed infection here). For simplicity, DNA transcription removed, a +ve. s.s. RNA virus is represented: its RNA acts directly as mRNA within the host system (vRNA → mRNA). The host ribosome is then used as efficient cause for viral proteins via folding and self-assembly. Critically, the material cause for these is necessarily supplied by the host cell. B. For comparison, diagram 30, of (Louie, 2013, Ch.11) redrawn to match the present symbol convention and with viral genes and their translated proteins (marked with prime) explicit (the original did not include primed labels, though the mapping stated with the diagram was given).

Showing that the virus is unable to independently achieve closure to efficient causation is much more than saying that they are obligate parasites because all organismal parasites are closed to efficient causation, only lacking some external (environmental) resources which they obtain from their host. Viruses, being essentially a linear chain of...
causal relations, provide no organisational demarcation between internal and external. In this respect they are no different from non-living things: they have no independence of agency, so lack the very essence of what it is to be alive. Without closure to efficient causation there is no life according to the organisational biology perspective within systems biology. In terms of causation, living things are definitively the efficient cause of themselves; efficient cause is necessarily the combination of formal and material cause; viruses are formal cause of themselves, but not material cause, so are not efficient cause and therefore cannot be living things.

Acknowledgement

During the review stage, Prof. J.H.S. Hofmeyr provided a lot of encouraging help and his contribution is acknowledged here. This paper was originally intended as a collaboration between the present author (KDF) and Prof. J. C. Letelier. However, in response to the COVID-19 medical emergency in his home country of Chile, Prof. Letelier dedicated all his efforts to developing a respiration support system which works and has presumably saved many lives. KDF would like to dedicate this work to the efforts of scientists (like Prof. Letelier) everywhere who have shelved their normal work to concentrate on saving lives during the 2020 pandemic. As a work of theory, this paper received no explicit financial support, but KDF acknowledges the generosity of his employer Queen’s University Belfast, in permitting time to develop and write it.

Footnote

I intend to use this informal section (not part of the published paper) to respond to scientific criticism of the work.

1. After reading this preprint, Dr Keith Baverstock challenged my statement that performing a thermodynamic work cycle meant the system must export entropy to its environment (line 134). He said “systems most emphatically do not export entropy.
A dissipative system increases in entropy and can gain order/complexity or maintain its structure”. He was of course referring to open systems and quoted Annila, A. and Baverstock K. (2016) Discourse on order vs. disorder, Communicative and Integrative Biology. V9. No. 4. e1187348 (open access). In response, during the proof-stage I added (more precisely, transform free energy) (line 135). I agree with him: entropy is not a thing, it is just a quantitative concept and what really happens is that the partition of energy into free and bound moves towards the bound by dissipation, for example high frequency photons are transformed into lower frequency ones.

2. I was contacted shortly after the publication of this article by microbiologist Claudiu Bandea. He is a leading proponent of the reductive theory on the origin of viruses: that they arose from an ancestral cellular parasite that lost a large number of genes over evolution, delegating much of the (costly) reproductive machinery to its host (see Bandea 2009). He refers to his theory of the origin of viruses as the fusion model, emphasising that originally intracellular parasites have fused with their host cells, leaving much of the original material and reproductive machinery redundant.

He put it to me that if viruses really are derived from formerly cellular (and organisationally autonomous) organisms, then it is hard to maintain that they are not now alive. In my view, there are no reasonable grounds to reject the reductive (fusion) theory, and I also accept the empirical support for it. However, I cannot think of a logical contradiction in saying a lineage could evolve out of life, if as my paper suggests, we define life by organisational closure. That is, viruses could have become so efficient (stripped down) as parasites that they no longer qualify as life using that definition (see section 4.2).

Claudiu challenged me to explain how a living parasitic cellular lineage suddenly evolved into a non-living parasitic viral lineage that continued to evolve and diversify into a myriad of other non-living viral lineages. The answer, I think is that the definition of living that I use does not preclude evolution and diversification. Evolution by natural
selection requires only variation (which is not an exclusive property of life), together
with competitive replication (which is evident in e.g. non-living chemical systems). A
complicated molecular system, that is a stripped down remnant of a cellular parasite
may evolve and diversify by replication within the host-cell environment without being
alive in the sense that I adhere to.

Of course, that depends on one’s definition of life, but those who prefer to define it
as a system capable of evolution by natural selection need to be sure they are happy to
include e.g. computer viruses and memes, or if they specify physical entities, the auto-
catalytic chemical systems I mentioned in the introduction. My main purpose in this
paper was to develop and offer a definition of life that transcends the known problems
with the others. That definition rests on organisational closure. For a definition to be
useful, it has to exclude some systems and, unfortunately for some virus researchers)
this definition excludes viruses. Despite that, I want to make it clear that viruses are
very intimately and crucially involved in life, almost certainly derived from life and, in
my view, may well be the remnants of once living organisms. I do recommend interested
readers to follow the published development of Claudiu Bandeas ideas and the strong
empirical support they have gained.

Bandeas CI. The Origin and Evolution of Viruses as Molecular Organisms. Nature
References

Abergel, C., Legendre, M., Claverie, J.M., 2015. The rapidly expanding universe of giant
viruses: Mimivirus, pandoravirus, pithovirus and mollivirus. FEMS Microbiology
Reviews 39, 779–796.

Abrahão, J., et al., S.L., 2018. Tailed giant tupanvirus possesses the most
complete translational apparatus of the known virosphere. Nat. Commun. 9,
https://doi.org/10.1038/s41467–018–03168–1.

grated causal structures in animats exposed to environments of increasing complexity.

systems - From elementary cellular automata to adapting animats. Entropy 17, 5472–
5502.

Auletta, G., Ellis, G., Jaeger, L., 2008. Top-down causation by information control:
from a philosophical problem to a scientific research programme. J. R. Soc. Interface


Thinking.. Springer, New York, USA.. chapter The problem of functional boundaries


Racaniello, V.R., 2014. Are viruses alive. URL: www.virology.ws/are-viruses-alive


54
van Regenmortel, M.H.V., 2016. The metaphor that viruses are living is alive and well, but it is no more than a metaphor. Studies in History and Philosophy of Biological and Biomedical Sciences 59, 117–124.


title{volume 13. chapter What is systems biology? From genes to function and back.}

A. 87, 4576–4579.

6, 643.

University Press, Oxford, UK.

Encyclopedia of Philosophy (Winter 2016 Edition). Metaphysics Research Lab, Stan-
mani.


Appendix - Details of IIT calculations

Analysis was conducted using the PyPhi package (python code) Mayner et al. (2018), downloaded via www.integratedinformationtheory.org.

A virocell (virus infected cell) was represented by adding the vRNA link into R and the vE link into E, with all viral contributions set ON. Modelling the cell alone (uninfected) was represented the same, but with the viral contributions set OFF.

1. Analysis of the full system (shown in Fig. 6)

The state transition matrix and connectivity matrix are given in Fig. 8 and Fig. 9.
Figure 8: The state transition matrix defining the full system cell Boolean network, used in IIT calculations.
Figure 9: The connectivity matrix defining the full system cell Boolean network, used in IIT calculations.
a: All nodes initially OFF: $\Phi = 0.125$.

- Concepts: $\varphi(R) = 0.25$; $\varphi(B) = 0.168$; $\varphi(P) = 0.168$; $\varphi(C) = 0.5$; $\varphi(E) = 0.5$.

b: All nodes initially ON: $\Phi = 0.3125$.

- Concepts: $\varphi(R) = 0.25$; $\varphi(B) = 0.25$; $\varphi(P) = 0.5$; $\varphi(C) = 0.5$; $\varphi(E) = 0.5$.

Cell only: The presence of the virus contributions made no difference to any part or whole of the IIT calculations.

2. Analysis of the Reduced System

Figure 10: The reduced cell Boolean network, with its connectivity and transition matrices used in IIT calculations.
All nodes initially OFF: $\Phi = 0.2153$.

- Concepts: $\varphi(P) = 0.168; \varphi(C) = 0.5$.

All nodes initially ON: $\Phi = 0.4375$.

- Concepts: $\varphi(P) = 0.5; \varphi(C) = 0.5$.

Again, including the virus as by setting $vRNA$ and $vE$ ON produced identical results to those obtained with the virus contributions OFF.

To find the attractors from the state transition matrices, one simply follows the trajectory reading $t \rightarrow t + 1$, setting the $t + 1$ state as $t$ and repeating until a pattern is recognised: this is done for every initial state. For both the complete and reduced system there are two point attractors: $\{0,0,0,0,0\}$; $\{1,1,1,1,1\}$ and $\{0,0,0\}$; $\{1,1,1\}$, respectively.

It is possible for anyone with the PyPhi package Mayner et al. (2018) to enter the state transition and connectivity tables given here to repeat the results.