JOURNAL OF THE ROYAL SOCIETY Interface

rsif.royalsocietypublishing.org

Review

Cite this article: Walker SI, Davies PCW. 2013 The algorithmic origins of life. J R Soc Interface 10: 20120869.
http://dx.doi.org/10.1098/rsif.2012.0869

Received: 23 October 2012
Accepted: 21 November 2012

Subject Areas:
biocomplexity, biophysics, astrobiology

Keywords:
origins of life, emergence, astrobiology, top-down causation

Author for correspondence:
Sara Imari Walker
e-mail: sara.i.walker@asu.edu

Royal Society Publishing
Informing the science of the future

The algorithmic origins of life

Sara Imari Walker[1,2,3] and Paul C. W. Davies[2]

[1]NASA Astrobiology Institute, Mountain View, CA, USA
[2]BEYOND: Center for Fundamental Concepts in Science, Arizona State University, Tempe, AZ, USA
[3]Blue Marble Space Institute of Science, Seattle, WA, USA

Although it has been notoriously difficult to pin down precisely what is it that makes life so distinctive and remarkable, there is general agreement that its informational aspect is one key property, perhaps the key property. The unique informational narrative of living systems suggests that life may be characterized by context-dependent causal influences, and, in particular, that top-down (or downward) causation—where higher levels influence and constrain the dynamics of lower levels in organizational hierarchies—may be a major contributor to the hierarchal structure of living systems. Here, we propose that the emergence of life may correspond to a physical transition associated with a shift in the causal structure, where information gains direct and context-dependent causal efficacy over the matter in which it is instantiated. Such a transition may be akin to more traditional physical transitions (e.g. thermodynamic phase transitions), with the crucial distinction that determining which phase (non-life or life) a given system is in requires dynamical information and therefore can only be inferred by identifying causal architecture. We discuss some novel research directions based on this hypothesis, including potential measures of such a transition that may be amenable to laboratory study, and how the proposed mechanism corresponds to the onset of the unique mode of (algorithmic) information processing characteristic of living systems.

1. Introduction

Of the many open questions surrounding how life emerges from non-life, perhaps the most challenging is the vast gulf between complex chemistry and the simplest biology: even the smallest mycoplasma is immeasurably more complex than any chemical reaction network we might engineer in the laboratory with current technology. The chemist George Whitesides, for example, has stated, 'How remarkable is life? The answer is: very. Those of us who deal in networks of chemical reactions know of nothing like it' [1]. The heart of the issue is that we do not know whether the living state is 'just' very complex chemistry, or whether there is something fundamentally distinct about living matter. Right at the outset, we therefore face a deep conceptual problem, one asked long ago by the physicist Erwin Schrödinger [2], namely, *What is life?* Without a definition for life, the problem of how life began is not well posed.

Often the issue of defining life is sidestepped by assuming that if one can build a simple chemical system capable of Darwinian evolution, then the rest will follow suit and the problem of life's origin will de facto be solved [3]. Although few are willing to accept a simple self-replicating molecule as living, the assumption is that after a sufficiently long period of Darwinian evolution this humble replicator will eventually be transformed into an entity complex enough that it is indisputably living [4]. Darwinian evolution applies to everything from simple software programs, molecular replicators and memes, to systems as complex as multicellular life and even potentially the human brain [5]—therefore spanning a gamut of phenomena ranging from artificial systems, to simple chemistry, to highly complex biology. The power of the Darwinian paradigm is precisely its capacity to unify such diverse phenomena, particularly across the tree of life—all that is required are the well-defined processes of replication with variation and selection. However, this very generality

is also the greatest weakness of the paradigm as applied to the origin of life: it provides no means for distinguishing complex from simple, let alone life from non-life. This may explain Darwin's own reluctance to speculate on the subject, 'One might as well speculate about the origin of matter', he quipped.

Although it is notoriously hard to identify precisely what makes life so distinctive and remarkable [6–8], there is general agreement that its informational aspect is one key property, and perhaps the key property [9–12]. The manner in which information flows through and between cells and sub-cellular structures is quite unlike anything else observed in nature. If life is more than just complex chemistry, its unique informational management properties may be the crucial indicator of this distinction [13]. Unfortunately, the way that information operates in biology is not easily characterized [10,14]. While standard information-theoretic measures, such as Shannon information [15], have proved useful, biological information has an additional quality which may roughly be called 'functionality'—or 'contextuality'—that sets it apart from a collection of mere bits as characterized by its Shannon information content. The information content of DNA, for example, is usually defined by the Shannon (sequential) measure. However, the genome is only a small part of the story. DNA is not a blueprint for an organism:[1] no information is actively processed by DNA alone [17]. Rather, DNA is a (mostly) passive repository for transcription of stored data into RNA, some (but by no means all) of which goes on to be translated into proteins. The biologically relevant information stored in DNA therefore has very little to do with its specific chemical nature (beyond the fact that it is a digital linear polymer). The genetic material could just as easily be a different variety of nucleic acid (or a different molecule altogether), as recently experimentally confirmed [18]. It is the functionality of the expressed RNAs and proteins—not the bits—that is biologically important.

Functionality, however, is not a local property of a molecule [19]. For example, the functionality of expressed RNA and protein sequences is clearly context-dependent—only an exceedingly small subset of these molecules is causally efficacious (i.e. meaningful) in the larger biochemical network of a cell whose functioning is dependent on conditions such as salinity of the cytoplasm, pH, etc. That milieu includes other expressed proteins, RNAs, metabolites and a host of other molecules, the spatial distribution of which is crucial to their individual causal roles. *A priori*, it is not possible to determine which will be functional in a cell based on local structure and sequence information alone.[2] One is therefore left to conclude that the most important features of biological information (i.e. functionality) are decisively non-local, subject to informational control and feedback, so that the dynamical rules will generally change with time in a manner that is both a function of the current state and the history of the organism [20,21] (suggesting perhaps that even the concept of evolution itself may be in need of revision, see Goldenfeld & Woese [21,22] for an insightful discussion).

The central position of information in biology is not itself especially new or radical [9–11]. What is often sidestepped, however, is the fact that in biological systems information is not merely a way to label states, but a property of the system. To be explicit, biological information is distinctive because it possesses a type of causal efficacy [23,24]—it is the information that determines the current state and hence the dynamics (and therefore also the future state(s)).[3] In this paper, we postulate that it is the transition to context-dependent causation—mediated by the onset of information control—that is the key defining characteristic of life. We therefore identify the transition from non-life to life with a fundamental shift in the causal structure of the system, specifically a transition to a state in which algorithmic information gains direct, context-dependent, causal efficacy over matter. We now turn to the question of how all this came about. How did information first gain causal purchase over certain complex systems that we now call living organisms?

2. Information in the origin(s) of life: traditional approaches

A longstanding debate—often dubbed the chicken or the egg problem—is which came first, genetic heredity or metabolism [25,26]? A conundrum arises because neither can operate without the other in contemporary life, where the duality is manifested via the genome–proteome systems. The origin of life community has therefore tended to split into two camps, loosely labelled as 'genetics-first' and 'metabolism-first'. In informational language, genetics and metabolism may be unified under a common conceptual framework by regarding metabolism as a form of analogue information processing (to be explained below), to be contrasted with the digital information of genetics. In approaching this debate, a common source of confusion stems from the fact that molecules play three distinct roles: structural, informational and chemical. In terms of computer language, in living systems chemistry corresponds to hardware and information (e.g. genetic and epigenetic) to software [27]. The chicken-or-egg problem, as traditionally posed, thus amounts to a debate of whether analogue or digital hardware came first.

2.1. A digital origin for life

The 'genetics-first' paradigm, identifying a digital information repository as the most essential feature of the first living systems, is favoured by biological approaches to the origin of life, which extrapolate backward in time from the properties of modern organisms. A widely accepted resolution to the seemingly inextricable duality of genotype–phenotype is that the modern 'DNA–protein' world evolved from simpler precursor system involving only one major molecular species that played both the role of information carrier and of enzymatic catalyst. In modern organisms, RNA is a biochemical mediator, enabling the translation of DNA to protein. RNA is unique in that it can fill both roles, acting as both a genetic polymer and a biochemical catalyst, with novel expanded roles for functional RNAs continually being discovered. This has led to the popular 'RNA world' hypothesis, where all known life is posited to have descended from an ancestral population of organisms that utilized RNA as their sole major biopolymer prior to the advent of DNA and protein [28–32].

Despite the conceptual elegance of the RNA world, the hypothesis faces problems, primarily because of the immense challenge of synthesizing RNA nucleotides under plausible

prebiotic conditions and the susceptibility of RNA oligomers to degradation via hydrolysis [33–35]. Some of the chemical difficulties are alleviated if RNA was preceded by an alternative genetic polymer such as peptide nucleic acid [36] or threose nucleic acid [37] (for other examples of candidate primitive genetic polymers, see Eschenmoser [38]). In genetics-first origin of life scenarios, it has therefore been suggested that early life may have undergone a 'hardware upgrade' (or a succession of upgrades), eventually transitioning from a proto-RNA genetic polymer (or even an inorganic substrate [39,40]) into RNA-based biochemistry at a later stage in its evolutionary history. This system would then have undergone further hardware upgrades or 'genetic-takeovers' to arrive at the DNA–protein world we observe today [41].

However, beyond the chemical difficulties associated with synthesis and stability of primitive genetic polymers [42], there lies a deeper conceptual challenge within the 'digital-first' picture. As remarked above, the proteome, and in fact nearly all biochemical interactions in the cell, processes information in an analogue format, i.e. through chemical reactions which rely on continuous rates. For example, much of the information digitally stored in DNA must first be transcribed and translated before it becomes algorithmically meaningful in the context of the cell where it is then processed as analogue information through protein interaction networks. Focusing strictly on digital storage therefore neglects this critical aspect of how biological information is processed. As we discuss below, because of the organizational structure of systems capable of processing algorithmic (instructional) information, it is not at all clear that a monomolecular system, where a single polymer plays the role of catalyst and informational carrier, is even logically consistent with the organization of information flow in living systems, because there is no possibility of separating information storage from information processing (that being such a distinctive feature of modern life). As such, digital-first systems (as currently posed) represent a rather trivial form of information processing that fails to capture the logical structure of life as we know it.

2.2. An analogue origin for life

In contrast to models that rely on extrapolating backward in time from extant biology, approaches that move forward from what is known of the geochemical conditions on the primitive Earth typically favour an analogue format for the first living systems. In analogue chemical systems, information is contained in a continuously variable composition of an assembly of molecules rather than in a discrete string of digital bits. 'Metabolism-first' scenarios for the origin of life fall within this analogue framework, positing that early life was based on autocatalytic metabolic cycles that would have been constructed in a manner akin to how analogue computer systems are cabled together to execute a specific problem-solving task [43,44]. The appeal of such metabolism-first scenarios is that the chemical building blocks—ranging from lipids [45], to peptides [46–48], to iron–sulphide complexes [49,50]—are usually much easier to synthesize under abiotic conditions than any known candidate genetic polymer and would have therefore been much more abundant on the prebiotic Earth. The heritable information in this case typically consists of the compositional ratios of the molecules in the organized assemblies. Although it has been suggested that such 'composomes' might provide a primitive inheritance mechanism [51,52], it is not clear that they are evolvable, since compositional information tends to degrade over successive generations inhibiting the capacity for open-ended evolution [53] (see [54] for a recent discussion of how such systems could be evolvable if possessing excess mutual catalysis). Therefore, informational inheritance is not nearly as clear cut here as it is in the digital picture.

Additionally, in the analogue-first picture there exists a deeper issue of (re)programmability and with the difficulty of maintaining orthogonal (i.e. non-interacting and thus non-interfering) reactions in strictly analogue reaction networks. Analogue computers fell out of favour in the mid-twentieth century because of issues of universality—analogue devices, regardless of their structure—are much more difficult to engineer to solve broad categories of problem than their digital counterparts. As we discuss below, all known life achieves universality (at least in a limited sense) by using the digital sequence structure of informational polymers. Such universality would be exceedingly difficult to engineer in an analogue-only system given the challenges associated with building reaction networks where each (programmed) reaction is chemically orthogonal to all other reactions. Orthogonality is, by comparison, relatively easy to achieve with digitized switches. Control is therefore much easier to achieve in an analogue system with digital switches than in a solely analogue system. Taking all of these factors into account, it is clear that analogue-only systems are not capable of adaptation in the same way as living systems are. Modern life is a hybrid: digital memory and digital switches enable control over many (non-interfering) analogue states, and therefore enable adaptability to changing environmental conditions with the same basic toolkit. This is another way of stating, in informational terms, that analogue-only systems are not as versatile or as robust as analogue systems with digital information control and as such may probably have very limited evolutionary capacity [55].

3. Redefining the problem: an algorithmic origin for life

By the above considerations, it seems that digital or analogue alone is insufficient to provide a satisfactory account of the origin of life—not just on technical grounds, but for deep conceptual reasons. The former suffers from difficulties of prebiotic synthesis and due to fundamental limitations on how information can be processed in such scenarios (being trivial rather than non-trivial, more on this below); whereas the latter suffers from issues of reprogrammability, control and potentially long-term evolvability. This dilemma forms the crux of the chicken-or-egg problem cited above and suggests that focusing solely on the debate over chemical hardware may be limiting progress. An implicit assumption of these traditional approaches has been that, while information may be manifested in particular chemical structures (digital or analogue), it has no autonomy. As such, information—though widely acknowledged as a key hallmark of life—thus far, has played only a passive role in studies of life's emergence. Instead, hardware has dominated the

discussion, in accordance with the generally reductionist flavour of biology in recent decades, with its associated assumption that, ultimately, all life is nothing but chemistry.

However, as stressed above, a rigorous distinction between life and non-life is most likely to derive from the distinctive mode of information management and control displayed by living systems, i.e. that in biology information is causally efficacious. Both the traditional digital-first and analogue-first viewpoints neglect the active (algorithmic or instructional) and distributed nature of biological information. In our view, an explanation of life's origin is fundamentally incomplete in the absence of an account of how the unique causal role played by information in living systems first emerged. In other words, we need to explain the origin of both the hardware and software aspects of life, or the job is only half finished. Explaining the chemical substrate of life and claiming it as a solution to life's origin is like pointing to silicon and copper as an explanation for the goings-on inside a computer. It is this transition where one should expect to see a chemical system literally take-on 'a life of its own', characterized by informational dynamics which become decoupled from the dictates of local chemistry alone (while of course remaining fully consistent with those dictates). Thus, the famed chicken-or-egg problem (a solely hardware issue) is not the true sticking point. Rather, the puzzle lies with something fundamentally different, a problem of causal organization having to do with the separation of informational and mechanical aspects into parallel causal narratives. The real challenge of life's origin is thus to explain how instructional information control systems emerge naturally and spontaneously from mere molecular dynamics. It is this issue which we explore in the remainder of this paper.

4. Turing, von Neumann and undecidability in the origin of life

The instructional, or algorithmic, nature of biological information was long ago identified as a key property, and an early attempt to formalize it was made by von Neumann. He approached the problem by asking whether it was possible to build a machine that could construct any physical system, including itself. Identifying the parallels between biological systems, such as the human nervous system, and computers, and drawing inspiration from Turing's work on universal computation, von Neumann [56] sought a formalism that would include both natural and artificial systems. Turing showed that it was possible to build a device, now known as a universal Turing machine, which, given a sufficient amount of time, could output any computable function [57]. A Turing machine is a relatively simple hypothetical device, consisting of a machine and an unlimited memory capacity taking the form of an infinite tape marked out into squares, on each of which a symbol may be printed or erased, sequentially. A key feature of Turing machines is that both the state of the machine and the current symbol on the tape being read in, are necessary to determine the future evolution of the system. As such, the algorithm encoded on the tape plays a prominent role in the time evolution of the state of the machine. At least superficially, this appears to be very similar to the case presented by biological systems where the update rules change in response to information read-out from the current state (as we discuss below, both are an example of top-down causation via information control). However, it is not obvious exactly how Turing's very abstract formalism might map onto biological systems. This was the problem von Neumann wished to solve.

By analogy with Turing's universal machine, he therefore devised an abstraction called a universal-constructor (UC), a machine capable of taking materials from its host environment to build any possible physical structure (consistent with the available resources and the laws of physics) including itself. An important feature of UCs is that they operate on universality classes.[4] In principle, an UC is capable of constructing any object within a given universality class (including itself, if it is a member of the relevant class). An example of such a universality class relevant to biological systems is the set of all possible sequences composed of the natural set of 20 amino acids found in proteins. The relevant UC in this case is the translation machinery of modern life, including the ribosome and associated tRNAs along with an array of protein assistants.[5] This system can, in principle, construct any possible peptide sequence composed of the coded amino acids (with minor variations across the tree of life as to what constitutes a coded amino acid [58]).

The UC forms the foundation of von Neumann's theory on self-replicating automata. However, an UC is a mindless robot, and must be told very specifically exactly what to do in order build the correct object(s). It must therefore be programmed to construct specific things, and if it is to replicate then it must also be provided with a blueprint of itself.[6] However, as von Neumann recognized, implicit in this seemingly innocuous statement is a deep conceptual difficulty concerning the well-known paradoxes of self-reference [59,60]. To avoid an infinite regress, in which the blueprint of a self-replicating UC contains the blueprint which contains the blueprint . . . ad infinitum, von Neumann proposed that in the biological case the blueprint must play a dual role: it should not only contain instructions such as an algorithm,to make a certain kind of machine (e.g. the UC) but should also be blindly copied as a mere physical structure, without reference to the instructions its contains, and thus reference itself only indirectly. This dual hardware/software role mirrors precisely that played by DNA, where genes act both passively as physical structures to be copied, and are actively read-out as a source of algorithmic instructions. To implement this dualistic role, von Neumann appended a 'supervisory unit' to his automata whose task is to supervise which of these two roles the blueprint must play at a given time, thereby ensuring that the blueprint is treated both as an algorithm to be read-out and as a structure to be copied, depending on the context. In this manner, the organization of a von Neumann automaton ensures that instructions remain logically differentiated from their physical representation. To be functional over successive generations, a complete self-replicating automaton must therefore consist of three components: an UC, a (instructional) blueprint and a supervisory unit.

To a rough approximation, all known life contains these three components, which is particularly remarkable, given that von Neumann formulated his ideas before the discoveries of modern molecular biology, including the structure of DNA and the ribosome. From the insights provided by molecular biology over the past 50 years, we can now identify

that all known life functions in a manner akin to von Neumann automata, where DNA provides an (partial) algorithm, ribosomes act as the core of the UC and DNA polymerases (along with a suite of other molecular machinery) play the role of a supervisory unit [60,61].[7]

In spite of the striking similarities between an UC and modern life, there are some important differences. DNA does not contain a blueprint for building the entire cell, but instead contains only small parts of a much larger biological algorithm, which may be roughly described as the distributed 'top-down' control of an organism. The algorithm for building an organism is therefore not only stored in a linear digital sequence (tape), but also in the current state of the entire system (e.g. epigenetic factors such as the level of gene expression, post-translational modifications of proteins, methylation patterns, chromatin architecture, nucleosome distribution, cellular phenotype and environmental context). The algorithm itself is therefore highly delocalized, distributed inextricably throughout the very physical system whose dynamics it encodes. Moreover, although the ribosome provides a rough approximation for an UC (see endnote 5), universal construction in living cells requires a host of distributed mechanisms for reproducing an entire cell. Clearly, in an organism the algorithm cannot be decomposed and stored in simple sequential digital form to be read-out by an appropriate machine in the manner envisioned by Turing and von Neumann for their devices.

Although the elements of von Neumann's UC cannot be put in a one-to-one correspondence with a living organism, the UC does provide a key insight into the nature of life, by directing attention to the logical structure of information processing and control, and information flow in living systems.

4.1. Trivial versus non-trivial self-replication

Although von Neumann automata are self-replicators, their mode of replication is non-trivial in a fundamental, logical sense, and should be distinguished from trivial replicators such as crystals, viruses, computer viruses, non-enzymatic template replicators, lipid composomes and Penrose blocks [62]. Cast in the language of the previous section, trivial replicators process information strictly in the passive sense. Typically, they are characterized by building blocks which are not much simpler than the assembled object. Schrödinger recognized this key distinction in his take on *What is life?* when he postulated that the genetic material must be some sort of 'aperiodic crystal' [2]. Algorithmic information theory can make the foregoing distinction precise. The algorithmic information of a system or structure is defined to be the Shannon information contained in the shortest algorithm that can specify the system or structure as its output [63–65]. For example, a trivial replicator, such as a crystal, is one that may be specified by an algorithm containing far fewer bits than the system it describes. In contrast, a non-trivial replicator is algorithmically incompressible and requires an algorithm, or instruction set, of complexity comparable to the system it describes (or creates).

A vast logical divide exists between trivial and non-trivial replicators because the former is not explicitly programmed. Instead, trivial replicators rely strictly on the implicit physics (and chemistry) of the current environment to support replication. Therefore, only a limited set of objects within a given universality class is constructible. In other words, trivial self-replicating systems can only access one instructional mode, the one which the system is currently operating in, and as such are capable of only passive information handling. This stands in stark contrast to the case for non-trivial replicators, where any possible object within a given universality class (as defined above), including the UC, can be constructed if the UC is provided with an appropriate instruction. Non-trivial replicators in some sense harness the underlying laws of physics and chemistry to achieve a broader agenda (although of course adhering to the constraints imposed by physical law). As such, only non-trivial replicators process information in an active sense, enabling the possibility for the update rules to change in response to the current informational state of the system (and vice versa). Because of this fundamental distinction in how information is handled and processed, non-trivial and trivial replication are two logically and organizationally distinct possibilities for self-replicating physical systems. The challenge in explaining life's origin is to account for the transition between trivial and non-trivial replication, which entails more than a mere leap in complexity, but a reconfiguration of the entire logical organization of the system.

4.2. Algorithmic takeover

Although modern life is clearly representative of the class of non-trivial self-replicators, the majority of work on the origin of life has focused on the conceptually simpler case of trivial self-replication. This is not without good reason: the origin of translation—mediating what is known of the transition from trivial to non-trivial[8]—is notoriously difficult to pin down, amounting to an algorithmic takeover of information stored in one molecular species (nucleic acids) that becomes operable over another structurally and chemically very different species (peptides). The division of labour implicit in bimolecular life bestows one very obvious and distinctive advantage; it enables the instructions to be physically separated and stored away from the hardware that implements them. The 'arm's length' control implicit in this division is exercised via a software channel—encoded transactions using messengers and specialized bilingual agents[9] that identify, and are read by a system that can decode the instructions. Thus, the algorithm inhabits one molecular universe and its products inhabit another. We consider this separation to be one of the hallmarks of life.

Although trivial self-replicators can undergo Darwinian evolution [66,67], the lack of separation between algorithm and implementation implies that monomolecular systems are divided from known life by a logical and organizational chasm that cannot be crossed by mere complexification of passive hardware. In that respect, we regard the case of the RNA world as currently understood as falling short of being truly living. If primitive 'life' was strictly monomolecular, there would be no way to physically decouple information and control from the hardware it operates on, resulting in unreliable information protocols because of noisy information channels. For this rather deep reason, it may be that life had to be 'bimolecular' from the start.

We point out a curious philosophical implication of the algorithmic perspective: if the origin of life is identified with the transition from trivial to non-trivial information processing—e.g. from something akin to a Turing machine

capable of a single (or limited set of) computation(s) to a universal Turing machine capable of constructing any computable object (within a universality class)—then a precise point of transition from non-life to life may actually be undecidable in the logical sense. This would probably have very important philosophical implications, particularly in our interpretation of life as a predictable outcome of physical law.

5. The origin of life: a transition in causal architecture

We have argued that living and non-living matter differ fundamentally in the way information is organized and flows through the system: biological systems are distinctive because information manipulates the matter it is instantiated in. This leads to a very different, context-dependent, causal narrative—with causal influences running both up and down the hierarchy of structure of biological systems (i.e. from state to dynamical rules and dynamical rules to the state) [68–71]. In modern life, genes may be up- or downregulated by physical and chemical signals from the environment. For example, mechanical stresses on a cell may affect gene expression. Mechanotransduction, electrical transduction and chemical signal transduction, all well-studied biological processes, constitute examples of what philosophers term 'top-down causation', where the system as a whole exerts causal control over a subsystem (e.g. a gene) via a set of time-dependent constraints [23,72,73]. The onset of top-down information flow, perhaps in a manner akin to a phase transition, may serve as a more precise definition of life's origin than the 'separation of powers' discussed in §4.2. The origin of life may thus be identified when information gains top-down causal efficacy over the matter that instantiates it. Top-down causation has an extensive literature so will not be reviewed here [23,24,68–70,73–75].

We note, however, that there may be several different mechanisms for top-down causation, which come into play at different hierarchical scales in nature [24]. As we have presented it here, the key distinction between the origin of life and other 'emergent' transitions is the onset of distributed information control, enabling context-dependent causation, where an abstract and non-physical systemic entity (algorithmic information) effectively becomes a causal agent capable of manipulating its material substrate [23,24].

Although there is an extensive literature on top-down causation, particularly in biology, it has not been explicitly applied to the origin of life as such. The framework presented in this paper provides a well-defined definition for the transition to life, drawing on the top-down concept within an informational framework. Such a definition also addresses the vexed issue of what constitutes 'almost life'. This is essential for any theory that purports to chart a directional pathway from simple building blocks towards progressively more 'lifelike' states. It makes sense to try to explain life's origin only if it resulted from processes of moderately high probability, so that we can reasonably expect to give an account in terms of known science. It then follows from simple statistics that there will have been a large ensemble of systems proceeding down the pathway toward life, and no obvious reason why only one member successfully completed the journey. Ideally then, there should be a parameter, or more probably a set of parameters, to quantify progress towards life. The causal efficacy of distributed information control, discussed throughout this paper, provides a plausible candidate parameter that includes the possibility of identifying states of 'almost life'.

Walker *et al.* [76] have recently proposed, via a toy model, one possible candidate measure for transitions in causal structure in biological hierarchies, using transfer entropy to study the flow of information from local to global and from global to local scales in a lattice of coupled logistic maps. Non-trivial collective behaviour was observed to emerge each time the dominant direction of information flow shifted from bottom-up to top-down, indicating that top-down causation was in fact driving the emergence of collectives. The particular dynamical system investigated was designed to parallel a hallmark of many major evolutionary transitions—the emergence of higher-level reproducers from previously autonomous lower-level units [77]. In this framework, the origin of life would mark the first appearance of this reversal in causal structure, and as such is a unique transition in the physical realm (marking the transition from trivial to non-trivial information processing as discussed earlier). The utility of this approach is that it provides a clear definition of what one should look for: a transition from bottom-up to top-down causation and information flow.

The aforementioned simple model, while instructive, suffers from the fact that it cannot capture how algorithmic information alters the update rules, and thus the future state of the system. A possible refinement is provided by Tononi's [78] measure of the so-called integrated information φ, based on network topology. This definition effectively captures the information generated by the causal interactions of the sub-elements of a system, beyond that which is generated independently by its parts. It therefore provides a measure of the distributed information generated by the network as a whole as a result of its causal architecture. Integrated information (also called 'excess information') has recently been successfully applied to measure emergence in cellular automata under appropriate coarse-graining of the dynamics [79]. A version of the theory whereby φ is in turn treated as a dynamical variable that then may influence the underlying causal relations among sub-elements might provide a way of quantifying the causal efficacy of information in the context discussed throughout this paper.

6. Conclusions

We have presented a framework for understanding the origin of life as a transition in causal structure, and information management and control, whereby information gains causal efficacy over the matter it is instantiated in. The hallmarks of living systems based on this approach as discussed in this paper are summarized in table 1. The advantage of this perspective is that it provides a foundation for identifying the origin of life as a well-defined transition. In so doing, it forces new thinking in how life might have arisen on a lifeless planet, by shifting emphasis to the origins of information control, rather than, for example, the onset of Darwinian evolution or the appearance of autocatalytic sets (i.e. either analogue or digital systems that lack information control), which, although certainly important to the story of life's

Table 1. The hallmarks of life.

hallmarks of life
global organization
information as a causal agency
top-down causation
analogue and digital information processing
laws and states coevolve
logical structure of a UC
dual hardware and software roles of genetic material
non-trivial replication
physical separation of instructions (algorithms) from the mechanism that implements them

emergence, do not rigorously define how/when life emerges as a function of chemical complexity. It also permits a broader view of life, where the same underlying principles would permit understanding of living systems instantiated in different chemical substrates (including potentially non-organic substrates). How this transition occurs remains an open question. While we have stressed that Darwinian evolution lacks a capacity to elucidate the physical mechanisms underlying the transition from non-life to life or to distinguish non-living from living, evolution of some sort must still drive this transition (even if it does not define it). It is probable that non-trivial information processing systems with delocalized information are more evolutionarily robust given that information can be preserved in the face of changing environmental conditions due the physical separation of information and its material representation.

Purely analogue life forms could have existed in the past but are not likely to survive over geological timescales without acquiring explicitly digitized informational protocols. Therefore, life forms that 'go digital' may be the only systems that survive in the long-run and are thus the only remaining product of the processes that led to life. As such, the onset of Darwinian evolution in a chemical system was probably not the critical step in the emergence of life. As we have discussed, trivially self-replicating systems can accomplish this. Instead, the emergence of life was probably marked by a transition in information processing capabilities. This transition should be characterized by a reversal in the causal flow of information from bottom-up only to a situation characterized by bi-directional causality. Methods to advance this programme include identifying the causal architecture of known biochemical networks by applying candidate measures (such as φ, or other measures of causal architecture [80,81]), and focusing on regulatory networks (information control networks) in ancient biochemical pathways to identify the minimal network architectures necessary to support the causal and informational narrative observed in extant life. A major unsolved problem is to determine how information control emerges ab initio, for example, in an RNA world setting, from chemical kinetics, as well as how primitive control mechanisms might evolve and become increasingly refined after 'algorithmic takeover' has occurred. Digitization may have been a natural outcome of this process in reaction networks that had once been primarily analogue. At this point, information would have become separated from its physical representation, permitting information to become a causal influence in its own right, and the language of Turing and von Neumann would have begun to apply. Characterizing the emergence of life as a shift in causal structure due to information gaining causal efficacy over matter marks the origin of life as a unique transition in the physical realm. It distinguishes non-living dynamical systems, which display trivial information processing only, from living systems (and the complex systems derivative of biological systems, such as computers) which display non-trivial information processing as two logically and organizationally distinct kinds of dynamical systems.

S.I.W. gratefully acknowledges support from the NASA Astrobiology Institute through the NASA Postdoctoral Fellowship Programme. S.I.W. also thanks the hospitality of the Aspen Center for Physics, supported in part by the National Science Foundation under grant no. PHY-1066293. P.C.W.D. was supported by NIH grant no. U54 CA143682. We thank Andrew Briggs, Luis Cisneros, John Doyle and George Ellis for stimulating conversations as well as the manuscript's anonymous reviewers for constructive comments.

Endnotes

[1]Whereas a blueprint provides a one-to-one correspondence between the symbolic representation and the actual object it describes, DNA does not contain all of the information necessary to reconstruct an organism [16]. For example, many post-translational modifications as well as self-assembling components (i.e. lipids) are not encoded in the genome.

[2]While some algorithms are becoming efficient at predicting structure, biological functionality is always determined by insertion in a cell, or inferred by comparison to known structures.

[3]The question of whether a causal chain expressed in informational language at the system level can ultimately be reduced, at least in principle, to a mechanistic causal chain at the molecular level, is the subject of a longstanding debate, complicated by the fact that biological systems are always open. We make no attempt to engage this notorious philosophical topic here, because it is irrelevant for the present discussion whether information is in fact a fundamental causal agent (which would represent a radical departure from standard physics), or may be treated merely phenomenologically as an effective causal agent.

[4]Here we define a universality class as the set of all possible objects that can be made from a given set of building blocks.

[5]The mapping between extant life and a von Neumann automaton is rather loose. In particular, the relevant UC here (i.e. the ribosome) is not included in the universality class it operates on and it therefore does not directly construct itself. There are a host of distributed control mechanisms and self-assembly processes that contribute to the reproduction of an entire cell.

[6]Likewise, an UC can construct any other object within its universality class if fed the appropriate instruction to do so.

[7]The all-important dual role cited earlier is clearly implemented: DNA polymerases are oblivious to the instructions that DNA contains and will blindly copy both coding and non-coding sequences.

[8]The informational narrative of life clearly goes beyond translation. However, this is the one place in biology where we know universality (at least in a limited sense) has taken hold. A complete mapping of epigenetic factors will probably uncover other informational protocols at work in biological systems that may have some form of associated universality, and perhaps are even more primitive.

[9]'Bilingual' here means tRNA molecules that recognize both the four-letter alphabet of nucleic acids and the 20-letter alphabet of amino acids.

References

1. Whitesides G. 2004 The improbability of life. In *Fitness of the cosmos for life: biochemistry and fine-tuning* (eds JD Barrow, SJ Freeland, SC Morris, CL Harper), p. xiii. Cambridge, UK: Cambridge University Press.
2. Schrödinger E. 1944 *What is life?* Cambridge, UK: Cambridge University Press.
3. Joyce G. 2012 Bit by bit: the Darwinian basis of life. *PLoS Biol.* **10**, e1001323. (doi:10.1371/journal.pbio.1001323)
4. Joyce G. 2002 Booting up life. *Nature* **410**, 278–279. (doi:10.1038/420278a)
5. Fernando C, Szathmáry E, Husbands P. 2012 Selectionist and evolutionary approaches to brain function: a critical appraisal. *Front Comput. Neurosci.* **6**, 1–28. (doi:10.3389/fncom.2012.00024)
6. Cleland CE, Chyba CF. 2002 Defining life. *Orig. Life Evol. Biosph.* **32**, 387–393. (doi:10.1023/A:1020503324273)
7. Tirard S, Morange M, Lazcano A. 2010 The definition of life: a brief history of an elusive scientific endeavor. *Astrobiology* **10**, 1003–1009. (doi:10.1089/ast.2010.0535)
8. Benner S. 2010 Defining life. *Astrobiology* **10**, 1021–1030. (doi:10.1089/ast.2010.0524)
9. Szathmáry E. 1989 The integration of the earliest genetic information. *Trends. Ecol. Evol.* **4**, 200–204. (doi:10.1016/0169-5347(89)90073-6)
10. Küppers BO. 1990 *Information and the origin of life*. Cambridge, MA: MIT Press.
11. Yockey H. 2005 *Information theory, evolution, and the origin of life*. Cambridge, UK: Cambridge University Press.
12. Hazen RM, Griffin P, Carothers JM, Szostak JW. 2007 Functional information and the emergence of biocomplexity. *Proc. Natl Acad. Sci. USA* **104**, 8574–8581. (doi:10.1073/pnas.0701744104)
13. Nurse P. 2008 Life, logic and information. *Nature* **454**, 424–426; 313–331. (doi:10.1038/454424a)
14. Maynard Smith J. 2000 The concept of information in biology. *Phil. Sci.* **67**, 177–194. (doi:10.1086/392768)
15. Shannon CE. 1948 A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423.
16. Shea N. 2007 Representation in the genome and in other inheritance systems. *Biol. Phil.* **22**, 313–331. (doi:10.1007/s10539-006-9046-6)
17. Noble D. 2008 Genes and causation. *Phil. Trans. R. Soc. A* **366**, 3001–3015. (doi:10.1098/rsta.2008.0086)
18. Pinhero VB *et al*. 2012 Synthetic genetic polymers capable of heredity and evolution. *Science* **336**, 341–344. (doi:10.1126/science.1217622)
19. Auletta G. 2011 *Cognitive biology: dealing with information from bacteria to minds*. Oxford, UK: Oxford University Press.
20. Gould SJ, Lewontin RC. 1979 The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. Lond. B* **205**, 581–598. (doi:10.1098/rspb.1979.0086)
21. Goldenfeld N, Woese C. 2011 Life is physics: evolution as a collective phenomenon far from equilibrium. *Ann. Rev. Cond. Matt. Phys.* **2**, 375–399. (doi:10.1146/annurev-conmatphys-062910-140509)
22. Goldenfeld N, Woese C. 2007 Biology's next revolution. *Nature* **445**, 369. (doi:10.1038/445369a)
23. Auletta G, Ellis GFR, Jaeger L. 2008 Top-down causation by information control: from a philosophical problem to a scientific research programme. *J. R. Soc. Interface* **5**, 1159–1172. (doi:10.1098/rsif.2008.0018)
24. Ellis GFR. 2012 Top-down causation and emergence: some comments on mechanisms. *Interface Focus* **2**, 126–140. (doi:10.1098/rsfs.2011.0062)
25. Orgel LE. 1998 The origin of life a review of facts and speculations. *Trends Biochem. Sci.* **23**, 491–495. (doi:10.1016/S0968-0004(98)01300-0)
26. Lazcano A, Miller S. 1996 The origin and early evolution of life: prebiotic chemistry, the pre-RNA world, and time. *Cell* **85**, 793–798. (doi:10.1016/S0092-8674(00)81263-5)
27. Davies PCW. 1999 *The fifth miracle: the search for the origin and meaning of life*. New York, NY: Simon and Schuster.
28. Gilbert W. 1986 Origin of life: the RNA world. *Nature* **319**, 618. (doi:10.1038/319618a0)
29. Cech T. 1993 The efficiency and versatility of catalytic RNA: implications for an RNA world. *Gene* **135**, 33–36. (doi:10.1016/0378-1119(93)90046-6)
30. Joyce G. 2002 The antiquity of RNA-based evolution. *Nature* **418**, 214–221. (doi:10.1038/418214a)
31. Robertson MP, Joyce G. 2010 The origins of the RNA world. *Cold Spring Harb. Perspect. Biol.* **4**, a003608. (doi:10.1101/cshperspect.a003608)
32. Gesteland RF, Atkins JF (eds) 1993 *The RNA world*. Cambridge, MA: Cold Spring Harbor Labortory Press.
33. Levy M, Miller SL. 1998 The stability of the RNA bases: implications for the origin of life. *Proc. Natl Acad. Sci. USA* **95**, 7933–7938. (doi:10.1073/pnas.95.14.7933)
34. Shaprio R. 2000 A replicator was not involved in the origin of life. *Life* **49**, 173–176.
35. Sutherland JD. 2010 Ribonucleotides. *Cold Spring Harb. Perspect. Biol.* **2**, a005439. (doi:10.1101/cshperspect.a005439)
36. Cline DB (ed.) 1996 *Peptide nucleic acid* (*PNA*). *Implications for the origin of the genetic material and homochirality of life*. New York, MA: American Institute of Physics.
37. Orgel LE. 2000 A simpler nucleic acid. *Science* **290**, 1306–1307. (doi:10.1126/science.290.5495.1306)
38. Eschenmoser A. 2007 The search for the chemistry of life's origin. *Tetrahedron* **63**, 12 821–12 843. (doi:10.1016/j.tet.2007.10.012)
39. Cairns-Smith AG. 1982 *Genetic takeover and the mineral origins of life*. Cambridge, UK: Cambridge University Press.
40. Davies PCW. 2004 Does quantum mechanics play a non-trivial role in life?. *Biosystems* **78**, 69–79. (doi:10.1016/j.biosystems.2004.07.001)
41. Leu K *et al*. 2011 The prebiotic evolutionary advantage of transferring genetic information from RNA to DNA. *Nucleic Acids Res.* **39**, 8135–8147. (doi:10.1093/nar/gkr525)
42. Engelhart A, Hud NV. 2010 Primitive genetic polymers. In *Cold Spring Harb. Perspect. Biol.* **2**, 1–21. (doi:10.1101/cshperspect.a002196)
43. Dyson FJ. 1982 A model for the origin of life. *J. Mol. Evol.* **18**, 344–350. (doi:10.1007/BF01733901)
44. Kauffman S. 1993 *The origins of order: self-organization and selection in evolution*. Oxford, UK: Oxford University Press.
45. Segré D, Ben-Eli D, Deamer D, Lancet D. 2001 The lipid world. *Orig. Life Evol. Biosph.* **31**, 119–145. (doi:10.1023/A:1006746807104)
46. Huber C, Wächtershäuser G. 1998 Peptides by activation of amino acids with CO on (Ni,Fe)S surfaces: implications for the origin of life. *Science* **281**, 670–672. (doi:10.1126/science.281.5377.670)
47. Lee DH, Granja JR, Martinez JA, Severin K, Ghadiri MR. 1996 A self-replicating peptide. *Nature* **382**, 525–528. (doi:10.1038/382525a0)
48. Childers SW, Ni R, Mehta AK, Lynn DG. 2009 Peptide membranes in chemical evolution. *Curr. Opin. Chem. Biol.* **13**, 652–659. (doi:10.1016/j.cbpa.2009.09.027)
49. Wachtershauser G. 1992 Groundworks for an evolutionary biochemistry: the iron–sulphur world. *Prog. Biophys. Mol. Biol.* **58**, 85–201. (doi:10.1016/0079-6107(92)90022-X)
50. Russell MJ, Hall AJ. 1997 The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. *J. Geol. Soc.* **154**, 377–402. (doi:10.1144/gsjgs.154.3.0377)
51. Segré D, Lancet D. 1999 A statistical chemistry approach to the origin of life. *Biochem. Mol. Biol.* **12**, 382–397.
52. Segré D, Lancet D. 2000 Composing life. *EMBO Rep.* **1**, 217–222. (doi:10.1093/embo-reports/kvd063)
53. Vasas V, Szathmáry E, Santos M. 2010 Lack of evolvability in self-sustaining autocatalytic networks constraints metabolism-first scenarios for the origin of life. *Proc. Natl Acad. Sci. USA* **107**, 1470–1475. (doi:10.1073/pnas.0912628107)
54. Markovitch O, Lancet D. 2012 Excess mutual catalysis is required for effective evolvability. *Artif. Life* **18**, 243–266. (doi:10.1162/artl_a_00064)
55. Chandra FA, Buzi G, Doyle JC. 2011 Glycolytic oscillations and limits on robust efficiency. *Science* **333**, 187–192. (doi:10.1126/science.1200705)
56. von Neumann J. 1958 *The computer and the brain*. New Haven, CT: Yale University Press.
57. Turing A. 1937 On computable numbers, with an application to the entscheidungsproblem. *Proc. Lond. Math. Soc.* **42**, 230–265. (doi:10.1112/plms/s2-42.1.230)

58. Knight RD, Freeland SJ, Landweber L. 2001 Rewiring the keyboard: evolvability of the genetic code. *Nat. Rev. Genetics* **2**, 49–58. (doi:10.1038/35047500)
59. Hofstadter D. 1979 *Godel, Escher, Bach: an eternal golden braid*. New York, NY: Basic Books, Inc.
60. Poundstone W. 1985 *The recursive universe: cosmic complexity and the limits of scientific knowledge*. New York, NY: William Morrow and Company Inc.
61. Mange D, Sipper M. 1998 Von Neumann's quintessential message: genotype + ribotype = phenotype. *Artif. Life* **4**, 225–227. (doi:10.1162/106454698568558)
62. Penrose LS, Penrose R. 1957 A self-reproducing analogue. *Nature* **179**, 1183. (doi:10.1038/1791183a0)
63. Kolmogorov AN. 1965 Three approaches to the quantitative definition of information. *Probl. Inform. Transm.* **1**, 1–7.
64. Chaitin GJ. 1969 On the simplicity and speed of programs for computing infinite sets of natural numbers. *J. ACM* **16**, 407–422. (doi:10.1145/321526.321530)
65. Chaitin GJ. 1987 *Algorithmic information theory*. Cambridge, UK: Cambridge University Press.
66. Eigen M. 1971 Self-organization of matter and evolution of biological macromolecules. *Die Naturwissenchaften* **58**, 465–523. (doi:10.1007/BF00623322)
67. Eigen M, Schuster P. 1977 The hypercycle: a principle of natural self-organization. *Die Naturwissenchaften* **64**, 541–565. (doi:10.1007/BF00450633)
68. Campbell DT. 1974 Downward causation in hierarchically organised biological systems. In *Studies in the philosophy of biology: reduction and related problems* (eds FJ Ayala, T Dobzhansky), pp. 179–186. London, UK: Macmillan.
69. Campbell DT. 1990 Levels of organization, downward causation, and the selection-theory approach to evolutionary epistemoloty. In *Theories of the evolution of knowing* (eds G Greenber, E Tobach). T.C. Schneirla Conference Series, pp. 1–15. Mahurah, NJ: Lawrence Erlbaum Associates.
70. Noble D. 2012 A theory of biological relativity: no privileged level of causation. *Interface Focus* **2**, 55–65. (doi:10.1098/rsfs.2011.0067)
71. Davies PCW. 2012 The epigenome and top-down causation. *Interface Focus* **2**, 42–48. (doi:10.1098/rsfs.2011.0070)
72. Levin M. 2012 Morphogenetic fields in embryogenesis, regeneration, and cancer: non-local control of complex patterning. *BioSystems* **109**, 243–261. (doi:10.1016/j.biosystems.2012.04.005)
73. Davies PCW. 2006 The physics of downward causation. In *The re-emergence of emergence* (eds P Clayton, PCW Davies), pp. 35–52. Oxford, UK: Oxford University Press.
74. Ellis GFR. 2006 On the nature of emergent reality. In *The re-emergence of emergence* (eds P Clayton, PCW Davies), pp. 79–107. Oxford, UK: Oxford University Press.
75. Ellis GFR, Noble D, O'Connor T. 2011 Top-down causation: an integrating theme within and across the sciences? *Interface Focus* **2**, 1–3. (doi:10.1098/rsfs.2011.0110)
76. Walker SI, Cisneros L, Davies PCW. 2012 Evolutionary transitions and top-down causation. *Proc. Artif. Life* **XIII**, 283–290.
77. Szathmáry E, Maynard Smith J. 1995 The major evolutionary transitions. *Nature* **374**, 227–232. (doi:10.1038/374227a0)
78. Tononi G. 2004 An Information integration theory of consciousness. *BMC Neurosci.* **5**, 42. (doi:10.1186/1471-2202-5-42)
79. Balduzzi D. 2011 Detecting emergent processes in cellular automata with excess information. (http://arxiv:11050158v2)
80. Pearl J. 2000 *Causality*. Cambridge, UK: Cambridge University Press.
81. Crutchfield JP. 1994 The calculi of emergence: computation, dynamics and induction. *Physica D* **75**, 11–54. (doi:10.1016/0167-2789(94)90273-9)