DUBLIN CITY
UNIVERSITY
Ollscoil Chathair Bhaile Átha Cliath

The Holland α-Universes Revisited*

Barry McMullin

September 1991

School of Electronic Engineering

TECHNICAL REPORT: `bmcm9101`

Barry McMullin,
School Of Electronic Engineering,
Dublin City University,
Dublin 9,
IRELAND

Telephone: +353-1-704 5432
Fax: +353-1-704 5508
E-mail: `McMullin@EENG.DCU.IE`

This is a slightly revised and re-formatted version of a paper presented at the First European Conference on Artificial Life, Paris, 1991.

Abstract

The α-Universes were introduced by John Holland as minimal systems in which the spontaneous emergence of self-replicating entities might be studied. This paper reviews Holland's original work and presents key results from an extended empirical investigation. The results indicate that, contrary to the original expectations, in the specific universe studied (α_0) the question of spontaneous emergence does not arise—because the putatively self-replicating entities prove not to be viable, even when artificially introduced.

1 Introduction

The theory that life on Earth arose spontaneously may be challenged by calculations based on unbiased random agitation of some more or less plausible chemical soup, showing that the expected emergence time would be many orders of magnitude longer than the time known to have been available. Holland (1976) argued that such calculations are critically flawed because the trajectory of the system would not be "random" but would, in fact, be strongly *biased* toward the emergence of life-like entities.

It is difficult to establish this point analytically, due to the complexity of the dynamic behaviour of real chemical systems. Holland therefore introduced the α-Universes: these are a class of *artificial* systems, providing crude and highly simplified analogs of certain chemical processes assumed to be involved in the origin of life. It is claimed that these systems are sufficiently complex that they can support identifiably "life-like" entities—while still being sufficiently simple that closed form analysis is possible (at least of certain aspects of their behaviour). Holland described one specific α-Universe in detail: I refer to this as α_0.

Holland presented a quantitative analysis of the emergence time for "life" in α_0, based solely on (analogs of) random, unbiased, agitation. He then showed that, far from α_0 actually exhibiting such an unbiased process, simple forerunners of "living" entities could be expected to emerge in a vastly shorter period, and would strongly bias the subsequent trajectory. If this analysis is correct then it establishes the *principle* that any analysis which presupposes an unbiased process in the origin of life is seriously flawed. Furthermore, it would indicate a practical approach toward the realisation of *artificial life* (albeit in a very primitive form)—simply by implementing α_0.

Holland's paper was limited to theoretical analysis; however, he noted that empirical testing of his results would be practical, using a computer implementation of α_0. The present paper reports on the outcome of just such a program of empirical testing.

2 The Universe α_0

Loosely speaking, α_0 consists of a 1-dimensional space of discrete *cells* (somewhat akin to a cellular automaton). In general, Holland is not explicit as to the size of α_0 nor (on the assumption that it is bounded) as to the behaviour at the boundaries. For numerical work, he typically uses a "region" of 10^4 cells. The analysis and experiments described in this paper assume an unbounded (circular) geometry, generally with a total size of 2×10^3 cells.

Each cell is either empty, or holds an *atom* of some *element*. Matter is conserved (i.e. the total number of atoms of each element is constant). Atoms may move (or be moved) around in the space, in accordance with certain dynamics, or *operators*, to be described below. Atoms in adjacent cells may be *bonded* together. Bonded atoms are guaranteed to remain adjacent (just so long as they stay bonded—bonds form and decay as part of the dynamics of the system).

A sequence of adjacent atoms, delimited at both ends by one or more empty cells, is called a *structure*. Note that the atoms making up a structure are not necessarily bonded. A *complex* is some specified *set* of structures. The structures making up a complex need not, in general, have a definite spatial relationship; however, for the complex to exhibit interesting properties it is generally necessary that all the component structures be more or less "close" to each other.

The dynamic behaviour of the α-Universe is defined by two groups of operators: the *primitive* operators, and the *emergent* operators. All operators are defined to operate stochastically. The primitive operators are context *insensitive*—i.e. they apply to all matter throughout α_0 without regard to its sequential organisation. They are the abstract counterparts of diffusion and activation in real chemical systems. The emergent operators are context *sensitive*—i.e. their operation is sensitive to the sequential organisation of matter. In effect, certain structures, (should they arise) have special dynamic properties. They are termed "emergent" operators precisely because they are contingent on such special configurations—they "emerge" iff matter in some region "happens" (under the action of the primitive operators or otherwise) to adopt some such special configuration. These are the abstract counterparts of catalysts (particularly enzymes) in real (bio-)chemical systems.

In the study of real chemical systems it is of interest to seek an explanation of the properties and characteristics of catalysis in terms of more fundamental (atomic) interactions. However, for the particular analyses we wish to make of the α_0 dynamics, such a more fundamental explanation would be superfluous, and is not attempted. Instead, the properties of emergent operators are simply imposed by fiat.[1]

---

[1] This is, in itself, an unusual and interesting (metaphysical) position. α_0 might be said to be, in a perfectly definite

Holland takes (self-)replication as diagnostic of "life"; the dynamics of α_0 are such that certain complexes (should they arise) may exhibit primitive self-replicating behaviours.

3 Predictions

The dynamic behaviour of the universe α_0 is, of course, critically affected by the presence, or otherwise, of emergent operators. Holland's approach is to assume that, starting from a "random" configuration, there will (almost certainly) be an initial epoch in which the density of emergent operators will be negligible, and the dynamics of the system can be analysed, to a good approximation, purely by reference to the primitive operators. I term this the *Primitive Epoch.* It is further assumed that, at some subsequent stage, the effects of emergent operators may become significant, and any analysis of the ensuing development will require *both* primitive and emergent operators to be taken into account. I term this the *Emergent Epoch.*

Holland further assumes (implicitly) that the transition (if any) from Primitive to Emergent Epochs will be triggered by the emergence of (more or less effective) self-replicating complexes. The argument is as follows: if structures (including emergent operators) are simply being formed and broken up again "randomly" (i.e. by the unbiased actions of the primitive operators) it seems plausible that their density (and thus their dynamic effects) will be small; whereas, if some system of structures (i.e. a complex), could actively *replicate* itself, instead of relying on random actions of the primitive operators for its formation, then it could "quickly" achieve a high density and would therefore have a significant impact on the dynamic behaviour of the system.

If this picture is accurate then the critical question becomes: how long is the Primitive Epoch expected to last?

By definition, during the Primitive Epoch, structures are generated and broken down more or less independently of their particular organisation. The statistics of this process are stationary. It is therefore possible to estimate the expected time to spontaneous emergence of any specific combination of structures (complex). Holland derives an approximate expression (his Theorem 3) for this emergence time. This shows (as might be anticipated) that the expected emergence time is critically dependent on the total size (number of atoms) of the specified complex; in fact, emergence time increases more or less as an *exponential* function of the size of the complex.

This general result can be applied to estimate the duration of the Primitive Epoch *only* on the basis of some identification of a particular complex (or perhaps any one of some set of complexes) whose emergence would trigger the transition to the Emergent Epoch. Holland considers two distinct cases, discussed in the following two sections.

3.1 The `FullSR` Complex

Suppose firstly that the transition cannot occur until a "fully" self-replicating complex appears. This is a complex which incorporates a *complete* "genetic" description of itself; this description is then (separately) copied and decoded to realise a replication cycle. This conforms essentially to the abstract model of self-replication pioneered by Schrödinger (1944) and (later) von Neumann (1951, 1966), and subsequently confirmed as the general mechanism exploited in all modern terrestrial life forms.

Holland identifies what is, more or less, the *simplest* α_0 complex which would exhibit this kind of fully self-replicating behaviour. I denote this complex by the symbolic name `FullSR`. It consists of 8 distinct, interacting, structures, composed of a total of 60 atoms.[2] The expected emergence time for `FullSR` is calculated to be approximately 10^{43} time steps (in a region of size 10^4 cells). For all practical purposes then, `FullSR` will *never* spontaneously appear, and it would appear that the critics of the idea of spontaneous emergence of life are vindicated (at least in α_0).

3.2 The `PartSR` and `Seed` Complexes

Alternatively, it may be the case that some significantly simpler complex, not capable of full replication as described above, might still be capable of some kind of *partial* replication, which would be sufficient to dramatically alter the subsequent dynamics of α_0. In particular, if such partially self-replicating complexes could become established, this would skew the distribution of structures in such a

---

and non-mystical sense, *irreducible.* That is, the properties and behaviours of structures in α_0 are not reducible to properties or behaviours of their "constituent" atoms. This is a tangible model of irreducibility in the sense of, say, Rosen's "complex" systems (Rosen 1985) or Popper's Worlds 1, 2 and 3 (Popper and Eccles 1977).

[2] This apparently small size, compared to the smallest comparable terrestrial organism, is an artefact of the relatively complex behaviours of quite small structures (emergent operators), which have been *designed* into α_0 to facilitate the spontaneous emergence of self-replication, even in a relatively small region (of space and time).

way that fully self-replicating complexes could then emerge relatively quickly (effectively via a conventional Darwinian process).

Holland identifies a complex of just 3 structures, with a total of only 14 atoms, for which this is apparently the case; I denote this complex by the symbolic name `PartSR`. The complex `PartSR` is not fully self-replicating, but it *can* replicate partial fragments of itself; if these fragments achieve a high density then a fraction of them will be spontaneously augmented and combined to achieve complete replication. Holland presents calculations to show that the fecundity of `PartSR` is indeed sufficient to achieve a net population growth rate, once the complex is initially established.

Holland goes on to show that an even smaller complex (2 structures, a total of 11 atoms) can function as an effective "seed" or precursor for the emergence of `PartSR`. I refer to this smaller complex as `Seed`. If the `Seed` complex ever appears (spontaneously) then it should apparently generate a substantial population of `PartSR` complexes; this `PartSR` population should then be viable in its own right even after the `Seed` complex decays.

Following this argument then, a transition from the Primitive to the Emergent Epoch will be triggered by the spontaneous emergence of `Seed`. The expected emergence time for this is calculated to be approximately 4×10^9 timesteps (again in a region of size 10^4 cells). If this analysis is correct then it suggests that the spontaneous emergence of "life" is quite feasible in α_0. Indeed, this result now admits the possibility of empirical test; Holland gives the example that, if an α_0 time step could be implemented in, say, 1ms of real time, then the expected emergence time for `Seed` becomes about 125 (real) hours—compared to over 10^{30} (real) centuries for `FullSR`...

4 Implementation

There are certain difficulties in implementing α_0 precisely in the manner described by Holland. For example, Holland implies a bounded universe, insofar as certain operator definitions require the space to be sequentially scanned from "left" to "right"; the effect of such operators is then uniquely defined only if "left" and "right" boundaries exist. On the other hand, however, the detailed behaviour of operators at universe boundaries is not consistently defined. Furthermore, certain operators, if implemented literally according to Holland's description, would be extremely inefficient in computational terms. The implementation has therefore been designed to preserve *only* the critical properties of α_0—i.e. those properties which are actually utilised in Holland's analysis; within this constraint it has been heavily optimised for execution speed.

The α_0 implementation has been written entirely in the C language (conforming to the ANSI X3J11 standard). Original development was carried out on IBM PC compatible machines, running MS-DOS, and Turbo-C V2.0. Most experimental study was carried out on this same platform—which incidently explains why a universe size of 2×10^3 cells has typically been used, this being the largest which can be comfortably accommodated on this platform. The package has also been ported to a VAX/VMS platform (which can support much larger universe sizes). As far as possible, the package has been written to be "easily" portable; machine dependancies are encapsulated by conditional compilation. The source code comprises about 3700 lines.

Performance varies widely with the particular configuration of the universe, and the platform in use—but one example is as follows: on a 286-based PC (8MHz clock), typical execution time is of the order of 20ms real time per α_0 time step, for a universe of 2×10^3 cells, with only the primitive operators enabled.

5 Playing God

To recap, there are three substantive elements to Holland's predictions:

1. The `Seed` complex will (spontaneously) appear in a relatively short time.
2. Once the `Seed` complex *does* appear, a population of `PartSR` complexes will be established, and will maintain themselves.
3. Conventional Darwinian evolution can then optimise the replicating ability of the complexes quite quickly.

Of these, the first potentially requires a substantial amount of (real) time to test; the second can be easily tested (by "playing God"—directly inserting an instance of the `Seed` complex); and the third can be tested only when (or if) testing of the second has been successful (i.e. after prediction 2 has been verified). Therefore, testing concentrated, in the first instance, on prediction 2—whether the `Seed` complex can establish a viable population of `PartSR` complexes. I present results for three scenarios, each calling for a progressively higher level of "divine intervention" in trying to trigger this emergence of artificial "life".

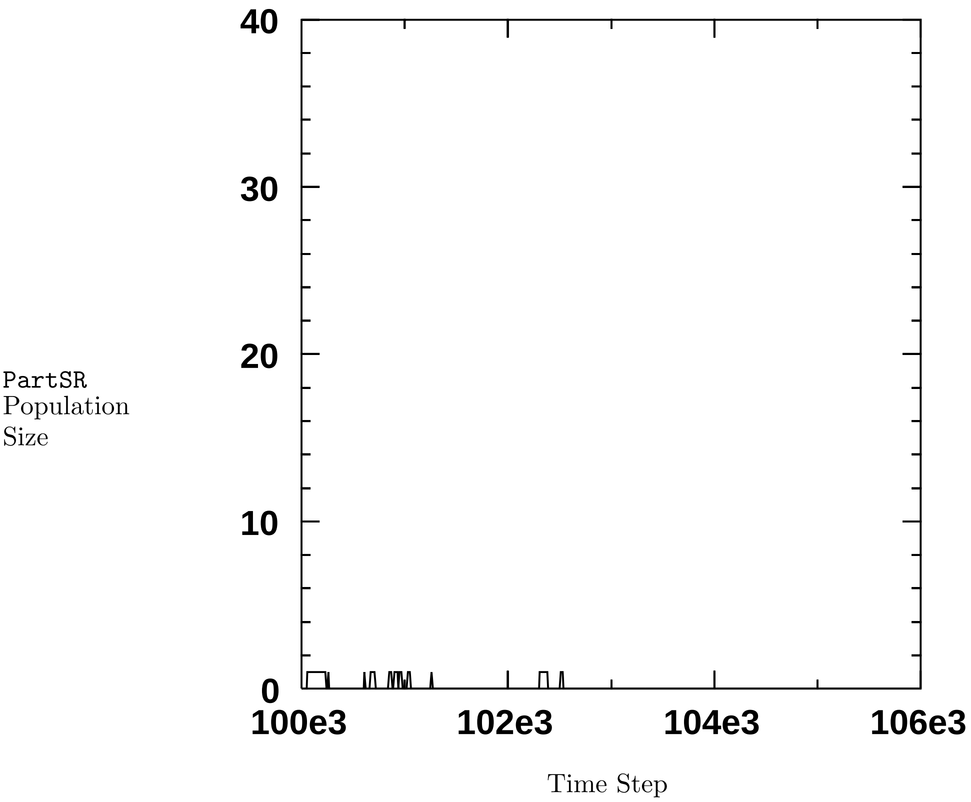

Figure 1: *"Economy" Model.* In this case, a randomised configuration is initially generated, transients are allowed to decay, and then a single instance of `Seed` is artificially inserted (at Time Step 100×10^3). Contrary to the original prediction, a large population of `PartSR` is *not* generated; in fact, the maximum `PartSR` "population" size achieved is just a single instance.

5.1 "Economy" Model

An α_0 of 2×10^3 cells was generated with a random initial configuration. This was executed for 100×10^3 time steps to allow transients, associated with the initial configuration, to decay. A single instance of `Seed` was then inserted. A further 6×10^3 steps were executed. Figure 1 is a plot of the number of `PartSR` complexes present over this latter period. It is clear that, contrary to expectations, a significant population of `PartSR` is *not* established (in fact, the maximum "population" achieved is just a single instance!).

5.2 "Standard" Model

It was noted that in normal operation of α_0 the density of free atoms (i.e. atoms with empty cells on both sides) was much lower than predicted by Holland's analysis, and generally approached zero. This has a detrimental effect on the operation of the `Seed` complex, as it relies on the availability of free atoms for its operation. It turns out that free atoms are generally not present because, even from a random configuration, there is a significant density of simple emergent operators. These bond any available free atoms together, more or less as they appear. In effect, Holland's assumption that there would *be* a Primitive Epoch, during which the action of the emergent operators could be neglected, is not valid.

However, it was not yet clear how serious this failing was. It was conjectured that the `Seed` complex might be able to operate effectively if the availability of free atoms was initially assured. Once a large population of the `PartSR` complex was established it could still then be viable (and support subsequent Darwinian evolution).

So an α_0 of 2×10^3 cells was again generated with a random initial configuration; but this time the single instance of `Seed` was *immediately* inserted. The initial random configuration has a significant density of free atoms (actually in accordance with the densities which Holland's analysis of the action of `Seed` were

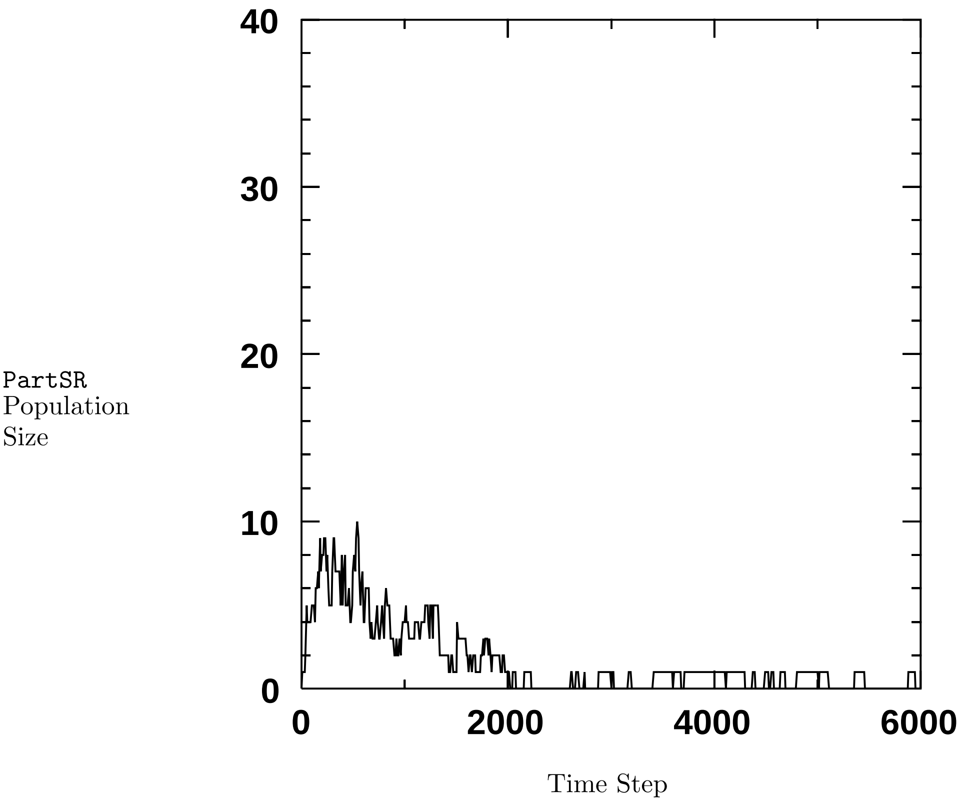

Figure 2: *"Standard" Model.* As in the "Economy" model, a randomised configuration is initially generated and a single instance of `Seed` is artificially inserted. However, in this case, `Seed` is inserted immediately (at Time Step 0)—i.e. without any delay to allow transients to decay; in particular, the (initial) density of free atoms is, in this case, relatively high. In contrast to the economy model, `Seed` does now succeed in generating a significant population of `PartSR` complexes—peaking at 10 instances, representing about 25% of the absolute maximum capacity for this particular universe. However, this population of `PartSR` is still evidently not viable in itself, and, contrary to prediction, rapidly goes extinct.

based on). 6×10^3 steps were executed. Figure 2 shows the resulting density of `PartSR`. The performance is certainly an improvement over the previous case, with the `PartSR` population reaching 10 instances; but the population still never approaches saturation (this particular universe has a capacity of 40 instances of the `PartSR` complex). Furthermore, after reaching a peak of about 10 instances, the population rapidly goes extinct again.

5.3 "Deluxe" Model

At this point it was clear that the `Seed` complex was not capable of carrying out the function anticipated by Holland—i.e. to establish a viable population of `PartSR` complexes. However, it was not clear whether this was merely a problem of the relatively limited size of `PartSR` population which the initial instance of `Seed` was managing to generate, or whether the `PartSR` complex would not be viable even in an arbitrarily large population.

To test this, an α_0 (still with 2×10^3 cells) was generated with a highly artificial initial configuration—namely saturated with instances of the `PartSR` complex. This was executed for 6×10^3 steps. Figure 3 is a plot of the number of `PartSR` complexes present over this period. It is seen that, even with this "most favourable" configuration, the population still rapidly goes extinct.

It was clear at this point that α_0 was not capable of supporting the "life-life" behaviour postulated by Holland. A number of variations were (briefly) investigated. These included increasing the size of the universe (to the 10^4 cells originally considered by Holland), and restricting the behaviour of the emergent operators so that, in effect, *only* the operators associated with `PartSR` would be executed. None

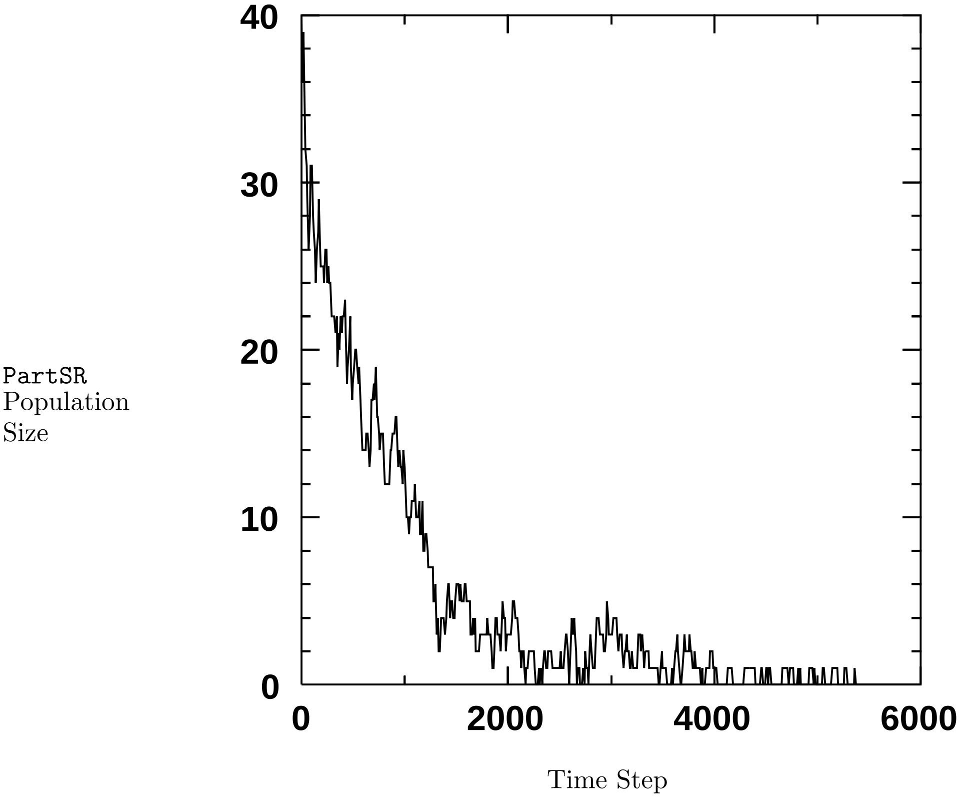

Figure 3: *"Deluxe" Model.* In this case a highly non-random initial configuration was generated—namely, saturated with instances of the `PartSR` complex (a total of 40 instances in this particular universe). It is seen that even in this "most favourable" case, which removes any reliance on the operation of the `Seed` complex, the population of `PartSR` is still not viable, and again rapidly goes extinct.

of these variations had a significant positive effect—the `PartSR` complex was not viable under any of the scenarios tested.

6 Analysis

It is not clear how deep rooted the deficiencies of α_0 are, but it is possible to identify some specific, proximate, causes of failure.

Holland's analysis relies on having a complex composed of structures which are, internally, "strongly bonded", which is to say *long lived.* He then estimates the average "productivity" over this lifetime, to come up with a net positive rate of change for the density of the `PartSR` complex (once a threshold is reached). However, in practise, there are (at least) three factors which severely disturb the behaviour of the complex, and which are not allowed for in Holland's analysis:

- Raw materials (free atoms) quickly become scarce (due to usage by random, garbage, emergent operators). This has already been discussed above; the effect is to drastically reduce the rate at which all emergent operators function in practice, thus reducing the *fecundity* of any putatively self-replicating complexes. However, given that the control experiments, in which this effect was artificially reduced or removed, still did not yield a viable `PartSR`, it seems that, though the effect is real, it is not critical (on its own).
- Even when a structure is internally bonded, there is nothing to stop random garbage moving into a position immediately adjacent to it. At the very least this interrupts or suspends the progress of an emergent operator. Thus, it turns out that complexes can only be *active* for a limited portion of their total lifetimes (regardless of the availability of free atoms); again this severely limits fecundity.

- But, at worst, this random arrival of a garbage structure beside an emergent operator can have much more severe effects. If it arrives on the right hand side it can corrupt the output of the operator (introducing a high "mutation" rate, and further reducing fecundity). If it arrives on the left hand side it can result in the formation of a different, garbage, emergent operator which forcibly, and prematurely, breaks up the original operator. This has actually been observed to occur on a number of occasions. Thus, as well as reduced fecundity, complexes also have higher mortality than expected.

So: compared to Holland's analysis, the lifetimes of the structures are shorter than expected, they are only active for a fraction of this time, and their products are sometimes corrupted. The net effect is that mortality exceeds fecundity (by a significant margin), the putatively replicating complexes cannot make up for their own natural decay rate, and thus become extinct quickly. These effects are directly related to the size of the complexes, and the time required to complete a replication cycle. Thus, they would actually affect the `FullSR` complex even more severely than `PartSR`.

A naïve attempt at solution of these problems might be to simply reduce the "temperature" of the universe (reduce the rate at which bonds decay, and structures get randomly moved around). It seems likely that the `FullSR` complex could be made viable in this way (in the limit, if the primitive operators are disabled entirely, `FullSR` would be able to expand to the capacity set by whatever free atoms are initially available; thereafter, of course, all dynamic activity would cease); however, this is not at all the case for `PartSR`, which relies on the primitive operators to complete its replication. It is quite possible (though no proof is currently available) that `PartSR` would not be viable at *any* "temperature". In any case, any reduction in "temperature" would be accompanied by an increase in the expected emergence time for any particular structure or complex, and thus may be completely counterproductive (given that spontaneous emergence was the problem originally being investigated).

The essential problem here seems to hinge on the fact that von Neumann style self-replication involves *copying* and *decoding* an information carrier, where the decoding must be such as to generate (at least) a copy of the required copying and decoding "machinery". α_0 fails to sustain this kind of behaviour because (*inter alia*) the maximum information capacity of its carriers (in the face of the various sources of disruption) seems to be of the order of perhaps 10 bits; this is insufficient to code for any worthwhile machinery—even the relatively simple copying and decoding machinery constructible in α_0.

A more plausible model for the spontaneous emergence of "life-like" behaviour may therefore involve a universe in which certain information carriers, of capacity (say) an order of magnitude larger than that required to code for minimal decoding machinery (in the particular universe), can be copied *without any specialised machinery at all.* In such a system there may be potential for a Darwinian evolutionary process to begin more or less immediately, in which more sophisticated phenotypic properties might, incrementally, become associated with the information carriers—possibly then culminating in a full blown "decoding" (or embryology).

This is, of course, rather speculative; but, as it happens, it is closely related to a general model for the origin of *terrestrial* life which has been championed by Cairns-Smith (1982). This is based on *inorganic* information carriers, which could conceivably be replicated without the relatively complex apparatus required for RNA or DNA replication. It seems to me, in the light of the experimental results presented here, that it would now be a promising research program to adopt Holland's original *strategy* (which is to design relatively simplified model chemistries, loosely based on cellular automata, in which to examine the origin of "life"), but to replace his detailed models (the α-Universes) with models based on different theoretical considerations—such as those of (for example) Cairns-Smith.

7 Conclusion

I should like to emphasise here the debt which this paper owes to John Holland's original formulation and analysis of the problem of spontaneous emergence of self-replicating behaviour. While it has been possible to point to defects in that analysis, this was with the benefit of hindsight, and prompted by experimental evidence not available to Holland. It does not detract in any way from Holland's creative achievement in formulating the *possibility* of such an investigation in the first place.

Drew McDermott once lamented that his field (Artificial Intelligence) was "starving for a few carefully documented failures" (McDermott 1985). It seems perverse to seek success *through* failure in this way—and such was certainly no part of my original intentions. Nonetheless, I think McDermott had a valid point, which is fairly well illustrated by the present work: for the primary conclusion here is that, although the model universe α_0 fails to demonstrate the phenomena originally hoped for, its particular mechanisms of failure are interesting and suggestive in their own right. In its own way, this may be best outcome of all.

Acknowledgments

This work was made possible by generous support from the School of Electronic Engineering in Dublin City University. I am also grateful to Noel Murphy of DCU, and John Kelly of University College Dublin, for many extended discussions relating to it, and to Inman Harvey of the University of Sussex for his perceptive criticism of an earlier draft.

References

Cairns-Smith, A.G. 1982. *Genetic takeover and the mineral origins of life*. Cambridge: Cambridge University Press.

Holland, J.H. 1976. Studies of the spontaneous emergence of self-replicating systems using cellular automata and formal grammars. In *Automata, Languages, Development*, edited by A. Lindenmayer and G. Rozenberg. New York: North-Holland.

McDermott, D. 1981. Artificial Intelligence Meets Natural Stupidity. In *Mind Design*, edited by John Haugeland. Cambridge: MIT Press.

Popper, K.R. and John C. Eccles. 1977. *The Self and Its Brain.* Berlin: Springer-Verlag.

Rosen, R. 1985. Organisms as causal systems which are not mechanisms: An essay into the nature of complexity. In: *Theoretical Biology and Complexity,* edited by Robert Rosen. Orlando: Academic Press.

Schrödinger, E. 1944. *What is Life?* Cambridge: Cambridge University Press.

Von Neumann, J. 1966. (Edited and completed by Arthur W. Burks) *Theory of Self-Reproducing Automata.* Urbana: University of Illinois Press.

Von Neumann, J. 1951. The general and logical theory of automata. In: *Cerebral Mechanisms in Behavior—The Hixon Symposium,* edited by L.A. Jeffress. New York: John Wiley. (Reprinted in Volume V of *John von Neumann—Collected Works,* edited by A.H. Taub. New York:Macmillan, 1961–63.)