John von Neumann and the Evolutionary Growth of Complexity: Looking Backwards, Looking Forwards...

Barry McMullin
http://www.eeng.dcu.ie/~mcmullin/

The final version of this article has been published in Artificial Life, Vol. 6, Issue 4, Fall 2000, pp. 347–361.
Artificial Life is published by The MIT Press.

Dublin City University
Research Institute for Networks and Communications Engineering
Artificial Life Laboratory

Abstract

In the late 1940's John von Neumann began to work on what he intended as a comprehensive "theory of [complex] automata". He started to develop a book length manuscript on the subject in 1952. However, he put this aside in 1953, apparently due to pressure of other work. Due to his tragically early death in 1957, he was never to return to it. The draft manuscript was eventually edited, and combined for publication with some related lecture transcripts, by Burks [2] in 1966. It is clear from the time and effort which von Neumann invested in it that he considered this to be a very significant and substantial piece of work. However: subsequent commentators (beginning even with Burks) have found it surprisingly difficult to articulate this substance. Indeed, it has since been suggested that von Neumann's results in this area are either trivial, or, at the very least, could have been achieved by much simpler means. It is an enigma. In this paper I review the history of this debate (briefly) and then present my own attempt at resolving the issue by focusing on an analysis of von Neumann's *problem situation* [16]. I claim that this reveals the true depth of von Neumann's achievement and influence on the subsequent deveopment of this field; and, further, that it generates a whole family of new consequent problems which can still serve to inform—if not actually define—the field of Artificial Life for many years to come.

1 Burks' Problem: Machine Self-Reproduction

> ... This result is obviously substantial, but to express its real force we must formulate it in such a way that it cannot be trivialized...
>
> —Burks [4, p. 49]

Arthur Burks makes this comment following a 46 page recapitulation of John von Neumann's design for a self-reproducing machine (realised as a configuration embedded in a 29-state, two dimensional, cellular automaton or CA). The comment, dating from 1970, can be fairly said to have initiated an extended debate about the significance of this work of von Neumann's, which has waxed and waned over time, but still persists to the present day.

Von Neumann's design is large and complex, and relies for its operation on exact and intricate interactions between the many relatively simple parts. In that sense, it is certainly substantial; but Burks is absolutely accurate in pointing out that this intricacy, in itself, is not necessarily interesting or significant. In particular, if the same "results" could be achieved, or the same "problem" solved, with drastically simpler machinery, then the interest of von Neumann's design would be critically undermined.

This is no idle concern on Burks' part. As he himself points out, within the CA framework, one can easily formulate a simple rule whereby a cell in a distinct state (labelled, say, `1`) will cause adjacent quiescent cells (labelled, say, `0`) to transit to state `1` also. By essentially the same definition or criterion as was applied to von Neumann's system, such a single cell, state `1`, configuration, would qualify as a self-reproducing machine—and would seem to render von Neumann's fabulously baroque design completely redundant.

Burks concluded that "... what is needed is a requirement that the self-reproducing automaton have some minimal complexity." And at first sight, this does seem eminently reasonable. Presumably (?) it is relatively easy to construct a "simple" machine, by whatever means; therefore it need not surprise us unduly if we manage to concoct a "simple" machine which can construct other "simple" machines, including ones "like" itself, and which therefore qualifies as self-reproducing. Whereas, it is relatively difficult to construct a "complex" machine, by any means; and therefore it may well be a challenging problem to exhibit a "complex" machine that is capable of self-reproduction. Von Neumann's machine certainly appears "complex", and certainly succeeds in constructing other machines like itself (i.e., in reproducing itself). So if we could just more formally express the precise sense in which von Neumann's machine *is* "complex", then we might indeed be able to clarify the "real force" of his achievement.

However, even at this point, we should be at least somewhat wary—because, while von Neumann himself certainly did introduce and discuss the notion of "complexity" in relation to this work, he did *not* attempt any formalisation of it. Indeed, he described his own concept of complexity as "vague, unscientific and imperfect" [23, p. 78]. It would therefore seem unlikely that the significance of his eventual results should actually *rely* on such a formalisation.

Nonetheless, Burks went on to propose just such a formal criterion of complexity—namely the ability to carry out universal computation. And by this criterion, von Neumann's design (or, at least, a straightforward derivative of it) would qualify as a "complex" self-reproducing

machine, and thus be clearly distinguished from the single cell, "1 state self-reproducer," which would remain merely a "simple" (and thus *trivial*) self-reproducing machine.

This seems a reasonable enough suggestion by Burks; though I would emphasise again that, as far as I have been able to establish, such a thing was never proposed by von Neumann himself—and, indeed, it jars seriously with von Neumann's calculated refusal to formalise complexity.

In any case, it turns out that Burks' suggestion is unsatisfactory and unsustainable. While it is true that von Neumann's machine (suitably formulated) can satisfy Burks' criterion for "complex" self-reproduction, this still represents an interesting result only if this criterion cannot be satisfied by very much simpler means. But in fact—and with hindsight, this now seems unsurprising—Burks' criterion *can* be satisfied by much simpler means than those deployed by von Neumann. This is because universal computation, *per se,* does not actually require particularly complex machinery [see, for example, 14].

This fact was formally demonstrated by Herman [8] in 1973, when he essentially showed how the basic single cell, 1 state, self-reproducer described earlier, could be combined with a single cell capable of universal computation. This results in a CA system in which the individual cells are "simpler" than in von Neumann's CA (i.e., have fewer states), and yet there are single cell configurations capable of both self-reproduction and universal computation. Granted, the universal computation ability relies on an adjacent, indefinitely long, "tape" configuration; but that was equally true of the universal computation ability of von Neumann's design, and is not a relevant distinguishing factor.

Herman draws the following conclusion [8, p. 62]:

> ... the existence of a self-reproducing universal computer-constructor in itself is not relevant to the problem of biological and machine self-reproduction. Hence, there is a need for new mathematical conditions to insure non-trivial self-reproduction.

So we see that Herman rejects Burks' specific criterion, while still continuing to accept Burks' formulation of the *issue* at stake—namely the identification of a suitable criterion for distinguishing "non-trivial" self-reproduction, albeit in Herman's version this is no longer explicitly tied to the notion of "complexity".

The discussion was taken up again by Langton [9] in 1984 (though apparently without reference to Herman's work). He presented a rather different analysis of Burks' criterion, but with ultimately complementary results. Langton pointed out that, as a general principle, there is little evidence to suggest that living organisms contain universal computation devices embedded within them. Since the self-reproduction of living organisms is presumably to be regarded as non-trivial (by definition—at least in this context), we should not, therefore, adopt universal computation as a criterion. So in this respect, albeit for different reasons, Langton concurs with Herman.

More importantly, Langton goes on to suggest a specific alternative criterion. He points out that self-reproduction in living organisms relies on a decomposition of the organism into two parts or components playing very distinct roles in the reproductive process:

1. The *genotype*, being an informational pattern stored or recorded in some sort of quasi-static or stable carrier. This information is transcribed or *copied* into a corresponding carrier in the offspring.

2. The *phenotype*, being the visible, active, dynamic, interacting part of the organism. The phenotype of the offspring is created by some sort of *decoding* or *interpretation* of the genotype (rather than by a copying of the parental phenotype).

Langton does not explain just *why* such a decomposition may be important, but rather seems to accept its pervasiveness among biological organisms as reason enough to adopt it as a criterion. And, in some sense it "works", because, indeed, von Neumann's self-reproducing design does have this architecture, whereas (say) Herman's self-declared "trivial" design does not. So it seems like this may be a satisfactory or even illuminating demarcation.

However: Langton did not stop there. With this new criterion in hand, he went on to consider whether it could be satisfied with a design significantly simpler than von Neumann's—and it transpires that it *can*. In fact, Langton was able to present a design for a CA space which itself is rather simpler than von Neumann's (i.e., having fewer states per cell), into which he could embed a self-reproducing automaton which, like von Neumann's, has an explicit decomposition into genotype and phenotype, but which is vastly smaller and simpler, occupying a region of just 150 cells—compared to the several hundred thousand cells required for von Neumann's device!

Langton's automaton is certainly still quite intricate, and its design involved a subtle interplay between designing the CA itself and designing the automaton to be embedded within it. In this sense it is a significant and substantive achievement. But is remains very unsatisfactory from the point of view of evaluating von Neumann's work. If Langton's criterion for non-trivial self-reproduction is accepted as an appropriate measure to judge von Neumann's work by, then we must still conclude that the latter's design is vastly more complex than necessary. While this *may* be true, I suggest that we should be reluctant to accept it without some much more substantive rationale for Langton's criterion. Or to put it another way, there may still be more "real force" to von Neumann's achievement than is captured or implied by Langton's criterion.

2 Von Neumann's Problem: The Evolutionary Growth of Complexity

I propose to resolve this enigma in a rather different way.

Firstly, I fully agree with Burks that to appreciate the full force of von Neumann's work, we must understand what *problem* he was attempting to solve. In particular, if it should turn out that this problem can be solved by trivial means (Herman), or at least much simpler means (Langton), then we should have to conclude that it was not such a substantial achievement after all. Where I differ from these, and indeed, most other, commentators, is that I think it is a mistake to view von Neumann's problem as having been wholly, or even largely, concerned with *self-reproduction*!

Of course, this is not to deny that von Neumann did, indeed, present a design for a self-reproducing automaton. I do not dispute that at all. Rather, my claim is that this self-reproducing capability, far from being the object of the design, is actually an incidental—indeed, *trivial*, though highly serendipitous—corollary of von Neumann's having solved at least some aspects of a far deeper problem.

This deeper problem is what I call the *evolutionary growth of complexity*. More specifically, the problem of how, in a general and open-ended way, machines can manage to construct other machines more "complex" that themselves. For if our best theories of biological evolution are correct, and assuming that biological organisms are, in some sense, "machines", then we must hold that such a constructive increase in complexity has happened not just once, but innumerable times in the course of phylogenetic evolution. Note that this claim does not rely on any sophisticated, much less formal, definition of complexity; it requires merely the crudest of qualitative rankings. Nor does it imply any *necessary* or consistent growth in complexity through evolution, but merely an acceptance that complexity has grown dramatically in *some* lineages.

Why is this growth of complexity a problem? Well, put simply, all our pragmatic experience of machines and engineering points in the opposite direction. In general, if we want to construct a machine of any given degree of complexity, we use even more complex machinery in its construction. While this is not definitive, it is surely suggestive of a difficulty.

To make all this a little more concrete, imagine that we could exhibit the following:

- Let there be a large (ideally, infinite) class of machine "types" or "designs"; call this M.
- Now consider those machine types, which we shall call *constructors*, which are capable of constructing *some* other (types of) machine. For any given $m \in M$ let $O(m)$ ("offspring of m") denote that subset of M which m is capable of constructing. m is then a constructor precisely provided $O(m) \neq \phi$; and—incidentally—m is self-reproducing provided $m \in O(m)$. This relation, $O(m)$, induces a directed graph on M; by following arrows on this graph we can potentially see what machine types can, directly or indirectly, be constructed by other machine types.
- Let there be a crude ("vague, unscientific and imperfect") measure of complexity on the elements of M—call it $c(m), m \in M$.
- Let $c(m), m \in M$ span a large (ideally, infinite) range, which is to say that there are relatively simple and relatively complex types of machine, and everything in between, included in M.

Now consider the two (highly schematic) graphs shown in Figure 1. In both cases I have shown the putative set M partitioned more or less coarsely by the complexity measure c. That is, the inner rings or subsets are those of small c (low complexity), while the outer, more inclusive rings or subsets include machines of progressively greater c (higher complexity). The graph on the left indicates our "typical" engineering experience: all constructional pathways lead inward (from more complex to less complex). As a result, complexity will always, and unconditionally, degenerate in time. Conversely, the graph on the right indicates the abstract situation posited by our best current theories of biological evolution: at least some edges of the graph (constructional pathways) lead from the inner rings to their outer neighbors. Provided there are sufficient such pathways, then, starting from only the very simplest machines (or organisms), there will be potential constructional pathways whereby

even the most complex of machines can eventually be constructed (in time). Thus complexity *might* grow in time.[1]

The problem which this presents is to show that the experience and intuition underlying the graph on the left is mistaken; to show, in other words, how the situation on the right might in fact be realised.

What would it take to solve this problem, even in principle? Well, one would need to exhibit a concrete class of machines M, in sufficient detail to satisfy ourselves that they *are* purely mechanistic; one would need to show that they span a significant range of complexity; and finally, one would have to demonstrate that there *are* constructional pathways leading from the simplest to the most complex (there may, or may not, also be pathways in the other direction, but that is not the point at issue—or at least, not immediately).

I believe that *this* is precisely the problem which von Neumann set out to solve. Furthermore, it seems to me that he did, indeed, solve it; and that it is only by seeing his work in *this* light that its "real force" can be understood.

3 Von Neumann's Solution

As to the first claim, that this—rather than "self-reproduction" per se—was von Neumann's problem, it is certainly relevant that he introduced his work in essentially these terms in his two documented public presentations of it—at the Hixon Symposium in September 1948 [24, p. 312] and at the University of Illinois in December 1949 [23, pp. 78–79]. Granted, in both cases he did *also* refer to the issue of self-reproduction. Specifically, he pointed out that self-reproduction is an example of a constructing machine where the output or offspring is just precisely matched in complexity to the parent—complexity neither increasing nor decreasing. But the critical point here—the significance of self-reproduction—is clearly as a special or "watershed" case of the more general situation: it is interesting precisely, and only, because it may mark a transition from strictly-degenerating to potentially-growing complexity. Conversely, if we encounter a form of "self-reproduction" that is *not* associated with such a transition, then it will not be of relevance to von Neumann's problem at all.

Indeed, von Neumann [23, p. 86] made this point even more explicitly, when he actually noted the triviality of self-reproduction in "growing crystals" (essentially Burks' 1-state reproducer in a CA). But *pace* Burks, *von Neumann's* resolution of this was not at all to impose a criterion of complexity, but rather to stipulate that the reproductive process should be such as to support "inheritable mutations". Taken in context, I think this must be interpreted as supporting at least some "mutations" where the offspring is of *increased* complexity—which again returns us squarely to the problem of the evolutionary growth of complexity, rather than self-reproduction.

My second claim—that von Neumann actually solved this problem—may seem more far fetched; but let us take it in steps.

First note that while von Neumann exhibited the design and operation of only one machine in detail—his self-reproducing machine—he consistently pointed out how his CA space

[1]Note that there are still inward, degenerative, pathways shown here; the claim is only that this graph permits the *possibility* of a growth in complexity; whether it actually *will* grow is a different question, and an altogether more difficult one. I will return to this briefly in the conclusion.

could support an indefinite variety of quite arbitrary machine configurations and behaviors. In this sense, he certainly exhibited not just a single particular machine, but a whole class of machines (which he identified via the infinite set of arbitrary, finite, "initially quiescent" configurations in his CA). This can correspond to the class M in my problem formulation above (though I will refine this somewhat below).

The next question is whether this class of machines spans a significant range of complexity. Given that we are using only the most vague definition of "complexity" here, the answer can be at best informal and qualitative. To my knowledge, von Neumann did not quite comment on this explicitly. He did explain at length that in setting up this sort of "axiomatic" theory of automata, there is necessarily a degree of arbitrariness in selecting the primitive or atomic parts from which the composite automata will be constructed. And indeed, between 1948 and 1953 he considered a number of alternative approaches before settling on a particular CA formulation (the latter concept having been suggested by Ulam—see [2, p. 94]). But given the highly complicated (co-ordinated, systematic, purposeful) behavior of the one machine he did design in detail, it certainly seems to me that if we allow, as von Neumann did, for the construction of machines with indefinitely more parts, in arbitrary configurations, then this set surely *might* span a sufficient range of complexity to meet our requirements.

This then leaves what seems by far the most intractable aspect of the problem: to show that there are constructional pathways leading from the simplest to the most complex of these machines. At first sight, it seems hopeless to demand a *demonstration* of this; firstly because the measure of complexity is so vague; but secondly because it would seem to demand separate analysis or demonstration of the constructional potentialities for most if not all of the elements of the infinite set M. But this, it seems to me, was precisely where von Neumann's crucial insight occurred.

Drawing on the capability of a single universal Turing machine to carry out the computations of any Turing machine at all, given a suitable description of that machine, von Neumann recognised that a single "general constructive automaton" [23, p. 85] might be able to construct "any" machine at all (i.e., any element of his specified M), given a "description tape" of that target machine. This immediately suggests the possibility that certain such special, "programmable", constructor (sub-)systems may enable enormous constructive potentialities—and also, incidentally (?) open up a powerful and *generic* route toward self-reproducing capability.

Von Neumann exploited this general idea by making a minor, but subtle, modification of his general constructive automaton (in a manner incidentally having no analog or significance in pure computation theory): as well as requiring it to *decode* a description to produce an offspring, he required that it must also *copy* the description, attaching this as part of the offspring. As Langton surmised, this combination of decoding and copying potentially has quite dramatic consequences—but the challenge is still to make those consequences explicit.

Let me denote a basic general constructive (decoding and copying) machine by u_0. We require that $u_0 \in M$.[2] Let $d(m), m \in M$ denote a "description of m", (relative to u_0). That is, letting $u_0 \oplus d(m)$ denote the composition of u_0 and a tape describing an arbitrary m, this will result in the construction of an instance of m, via *decoding* of the description, itself

[2]It is neither trivial nor obvious that there exists a general constructive automaton within any given M; this is why the bulk of von Neumann's unfinished manuscript is taken up with the detailed design of a particular example u_0 to establish this result for his particular M.

composed with a *copy* of the original tape, i.e., $m \oplus d(m)$. We write this as:

$$(u_0 \oplus d(m)) \rightsquigarrow (m \oplus d(m))$$

Since this applies for all $m \in M$ it applies to u_0 itself and we have:

$$(u_0 \oplus d(u_0)) \rightsquigarrow (u_0 \oplus d(u_0))$$

This is the single self-reproducing automaton, which has been commonly identified as von Neumann's "result" or "achievement". Certainly at this point we have a case of construction where complexity has been fully preserved—but the question remains whether this gives us an avenue into the case where complexity can grow.

Now note that u_0 can be combined or augmented with fairly arbitrary ancillary machinery, while still retaining its general constructive ability. That is, we can say that $u_0 \oplus m$ is still a general constructive automaton for "almost all" $m \in M$. The exception would be where m in some sense interferes with or disrupts the normal operation of u_0. Discounting those cases, we can say that the existence of a single general constructive automaton, u_0, actually implies the existence of a whole infinite set of related general constructive automata, $U \subset M$, where $u_m \in U$ has the form $u_m = (u_0 \oplus m)$ (i.e., all the members of this set share the same "general constructive sub-system"). This, in turn, implies the existence of a whole infinite set of self-reproductive automata, S where each $s_m \in S$ has the form $u_m \oplus d(u_m)$. Furthermore, given that U is derived from M just by excluding some exceptional or pathological cases (i.e., where the operation of u_0 would be disrupted), we can say that U, and thus S, will still span essentially the same significant range of complexity as M itself.

Now this is clearly a much stronger result than merely demonstrating the existence of a single self-reproducing machine. In particular, it indicates the possibility of arbitrarily complex machines that are still capable of self-reproduction. This in itself certainly distinguishes von Neumann's result from that of Langton [9]. Although the operation of Langton's machine can be decomposed into copying and decoding activities, it does not incorporate anything like a "general constructive automaton". The decoding capability is extremely impoverished so that, in effect, there is only one description that can be effectively processed—that of the one self-reproducing configuration which Langton exhibited.

But in any case, there is still a further critical step in the analysis of von Neumann's work. The goal is not just to show a class of arbitrarily complex constructors capable of constructing machines of equal complexity (which is what self-reproduction illustrates), but to demonstrate constructional pathways whereby complexity can *grow*. And it turns out that this further result can also be demonstrated by the set S.

To see this, we imagine the possibility of perturbations—"mutations"—to the description tape of a machine $s = (u_m \oplus d(u_m)), \in S$. With reasonable assumptions about the description language, it can be arranged that all, or almost all, tape configurations are "legal" (i.e., can be decoded to *some* target machine), and that the tape $d(u_m) = d(u_0 \oplus m)$ can be decomposed into a part, $d(u_0)$, coding for u_0, and a part, $d(m)$, coding for the ancillary machinery, m. That being the case, a more or less arbitrary modification of the $d(m)$ portion of the tape will result in a legal description of some *other* machine, $d(m')$. The reproductive cycle will then result, not in self-reproduction, but in a "mutant" offspring $s' = (u_{m'} \oplus d(u_{m'}))$; but this is still an element of S, and thus self-reproducing in its own right.[3]

[3]Indeed, we call such changes "mutations" precisely because they can "breed true" in the offspring.

Now these mutations essentially open up additional constructional pathways *between* the elements of S. In particular it is now clear that this can allow incremental, bootstrapping, *increases* in complexity. In effect, the density of these mutational pathways reflects the combinatorics of the description code, so that we can be virtually guaranteed, *without any detailed analysis,* that there will be constructional pathways connecting the entire set S. These will include, of course, degenerating pathways, where the offspring are less complex; but will also include vast numbers of pathways of increasing complexity.

In this way, finally, we see how von Neumann's detailed design of a single machine implies at least a schematic solution of the generic problem of the evolutionary growth of machine complexity.

4 Looking Backwards

While the designation of a distinct field known as *Artificial Life* is comparatively recent [10], I would argue that, looking backwards, it is clear that von Neumann's work properly defines its origin and inspiration. If the formulation above of von Neumann's problem—and his solution to it—is accepted, then a proper assessment of his contribution to the field he effectively founded presents at least three distinct questions:

1. Was von Neumann the *first* one to solve his problem?
2. Has von Neumann presented the *only* known solution to his problem?
3. Is von Neumann's solution the *simplest* known?

My answer to question 1 is clearly an unambiguous "yes"; and it stands as von Neumann's remarkable achievement *both* to have formulated this foundational problem *and* to have actually succeeded in solving it.

Question 2 is slightly more difficult. Given than von Neumann's problem situation has been poorly understood, it follows that subsequent contributors have not generally articulated clearly the relation between their work and this problem.

However, I do not hesitate to say that Thatcher [20], Codd [6] and Pesavento [15] have all offered alternative, or at least significantly refined, solutions. Thatcher provided a somewhat simplified design for the basic general constructive automaton u_0, in the *same* CA space as defined by von Neumann. Codd defined a significantly simpler CA space, which could still accommodate a functionally equivalent general constructive automaton. Much more recently, Pesavento has demonstrated a substantially simplified design for u_0 in a CA space marginally more complicated than von Neumann's; and, further, has actually demonstrated a functioning simulation of this complete u_0. Because these three systems are so very closely related to von Neumann's it follows that they are at least equally satisfactory in solving his problem.

Conversely, as already indicated, I think we can be clear that Langton's "self-replicating loop" system [9] does *not* qualify as an alternative solution to von Neumann's. Although it was derived from Codd's earlier system, and although Langton exhibits a self-reproducing configuration (involving "copying" and "decoding") this does not embody anything like a "general constructive automaton" and therefore has little or no evolutionary potential.

Another possible candidate solution to von Neumann's problem might be that of Berlekamp et al. [1]. There it is claimed that a general constructive automaton, of comparable functionality to von Neumann's u_0, can be realized in Conway's well known "Game of Life" (GOL) CA. This arguably takes Codd's (and thus von Neumann's) result to some kind of limiting case, as the GOL cells have the absolute minimum of only 2 states. On the other hand, in contrast to the fully detailed designs for u_0 presented by Thatcher, Codd and Pesavento, Berlekamp et. al. provide only a very sketchy outline to indicate the *possibility* of an equivalent automaton in GOL. Moreover, because GOL is so radically simplified (as a CA) it would not be quite trivial to establish that it supports embedded automata of comparable range of "complexity" to those of Von Neumann etc.

This last issue affects the comparative evaluation of certain other systems even more seriously. I have in mind particularly `Tierra` [17] and related systems.

`Tierra` is, roughly, a shared memory parallel computer, with fairly simple "von Neumann style" processors.[4] Ray has exhibited processes (embedded "machines" or "automata" in von Neumann's sense) in this framework which are capable of self-reproduction; and which, moreover, exhibit clear and significant evolutionary change. However, because the `Tierra` system is so very different in style from von Neumann's (or, indeed, *any* CA of 2 or more dimensions), it is very difficult to make even informal and qualitative comparisons of automata "complexity" between these systems.

I also have another quite different, and perhaps more important, misgiving about whether `Tierra` should be regarded as offering an equally satisfactory alternative solution to von Neumann's problem.

I originally expressed this reservation by stating that `Tierra` uses "... a form of self-reproduction based on self-inspection (rather than a properly genetic system in the von Neumann sense)" [12, Chapter 4]. The implication was that the self-reproducing entities in `Tierra` lack the important distinction between "copying" and "decoding"—and that this perhaps affects evolutionary potential in the system. However Taylor [19] has since persuaded me that this way of presenting the matter is, at best, unclear.

It might be more precise to say that `Tierra` *does* incorporate a distinction between copying and decoding processes: but that the "decoding" is represented by the execution of instructions by the `Tierra` processors. Thus, the decoding is hard-wired, whereas, in von Neumann's system, the decoding is itself a product of the specific design of u_0, and, as such, is at least *potentially* mutable. In this way, von Neumann's system allows for what I have called *Genetic Relativism* [12, Chapter 4], where the actual "decoding" map can itself evolve. It seems to me that this allows for a more profound form or degree of evolvability in von Neumann's system compared to `Tierra`.

Now this is an idea which von Neumann himself seems to have recognised, but discounted. He explicitly asserted that mutations affecting that part of a descriptor coding for u_0 would result in the production of "sterile" offspring [23, p. 86]—and would thus have no evolutionary potential at all. Clearly, on this specific point, I disagree with von Neumann, and consider that he actually underestimated the force of his own design in this particular respect. Accordingly, I will explore this issue in more depth in the following section below.

[4]Here, of course, I am referring to the so-called "von Neumann computer architecture" [22] (which was not, by any means, solely von Neumann's invention) rather than to his very different work on cellular automata.

As to my question 3 above—whether von Neumann's solution is the "simplest"—this clearly does not admit of a clearcut answer, given the informality of our notions of simplicity and complexity. Certainly, the alternative solutions of Thatcher, Codd and Pesavento can all be said to be *somewhat* simpler in at least some respects. Similarly, if the outline design of Berlekamp et. al. is accepted then it is *much* simpler in one particular dimension (the number of cell states). Nonetheless, in considering the overall development of the field I would say that none of these "simplifications" either undermine, or significantly extend, von Neumann's seminal results.

5 Evolvability: A Closer Look

The set of von Neumann self-reproducers anchored on a single u_0 have precisely this in common: they process the same formal "genetic language" for describing machines. In biological terms we may say that this set incorporates a fixed, or *absolute* mapping between genotype (description tape) and phenotype (self-reproducing automaton).

Thus, in committing ourselves (following von Neumann) to solving the problem of the evolutionary growth of complexity purely within the resources of a single such set, we are also committing ourselves to the equivalent of what I call *Genetic Absolutism* [13, Section 5.3]. I should note that, in the latter paper, I argue at length against the idea of Genetic Absolutism; but not in the sense that it is "bad" in itself—it just is not a tenable theory of biological evolution. Now von Neumann is not yet trying to capture all the complications of biological evolution: he is merely trying to establish that some key features, at least, can be recreated in a formal, or artificial, system. If this can be done within what is, in effect, a framework of Genetic Absolutism, *and if there is some advantage to doing this in that particular way,* then the fact that it is still "unbiological" (in this specific respect) should not be held too severely against it. (Indeed, there are arguably much more severe discrepancies than this in any case.)

Now, as it happens, adopting Genetic Absolutism *does* have a significant advantage for von Neumann. Working within such a framework it *is* necessary (for the solution of von Neumann's problem) to exhibit one core general constuctive automaton, u_0; and it *is* necessary to establish that this is sufficiently powerful to satisfy the informal requirements of the evolutionary growth of complexity; and it *is* finally necessary to show that, based on the formal genetic language processed by u_0, there is a reasonable likelihood that most, if not all, of the corresponding self-reproducers will be directly or indirectly connected under mutation. But if all this can be done, then the problem immediately at issue for von Neumann can, indeed be solved.

But the key point is that even though this may *suffice* for von Neumann's immediate purposes, nonetheless his framework is actually capable of going well beyond this; and I will claim that there may be advantages in doing so.

As I indicated in the previous section, the alternative to Genetic Absolutism is *Genetic Relativism* [13, Section 5.4], which envisages that the mapping between genotype (description tape) and phenotype (self-reproducing automaton) is *not* fixed or absolute but may vary from one organism (automaton) to another.

If we tackle von Neumann's problem in a framework of Genetic Relativism, we do *not*

restrict attention to a single u_0, giving rise to an "homogenous" set of self-reproducers, all sharing the same genetic language. Instead we introduce the possibility of having many *different* core automata—u_0^1, u_0^2, etc. Each of these will process a more or less *different* genetic language, and will thus give rise to its own unique set of related self-reproducers. We must still establish that most if not all self-reproducers in each such set are connected under mutation; but, *in addition,* we must try to show that there are at least some mutational connections between the *different* such sets (in order to establish pathways for evolution of the mapping itself).

The latter is, of course, a much more difficult task, because the mutations in question are now associated with changes in the very languages used to decode the description tapes. But, if such connections could be established, then, for the purposes of solving von Neumann's problem we are no longer restricted to considering the range of complexities of any single von Neumman set of self-reproducers (i.e., anchored on a single u_0, with a common description language), but can instead consider the union of many—indeed a potential infinity—of such sets.

Now clearly, in terms simply of solving von Neumann's problem, Genetic Relativism introduces severe complications which are not necessary, or even strictly useful. For now we have to exhibit not one, but multiple core general constructive automata, processing not one, but multiple genetic languages; and we have to characterise the range of complexity, and mutational connectivity, of not one but multiple sets of self-reproducers; and finally, we still have to establish the existence of mutational links *between* these different sets of self-reproducers. At face value, the only benefit in this approach seems to be the rather weak one that maybe—just maybe—the distinct general constructive automata can be, individually, significantly simpler or less powerful than the single one required under Genetic Absolutism; but it seems quite unlikely that this could outweigh the additional complications.

Let me say then that I actually accept all this: that for the solution of von Neumann's problem, as I have stated it, adopting the framework of Genetic Absolutism seems to be quite the simplest and most efficacious approach, and I endorse it as such. Nonetheless, I think it worthwhile to point out the *possibility* of working in the alternative framework of Genetic Relativism for a number of distinct reasons.

Firstly, it would be easy, otherwise, to mistake what is merely a pragmatic preference for using Genetic Absolutism in solving von Neumann's problem with the minimum of effort, for a claim that Genetic Absolutism is, in some sense, *necessary* for the solution of this problem. It is not. More generally, our chosen problem is *only* concerned with what may be possible, or sufficient—not what is necessary.

A second closely related point is this: *prima facie,* our solution based on Genetic Absolutism may seem to imply that a *general* constructive automaton (i.e., capable of constructing a very wide range of target machines) is a pre-requisite to *any* evolutionary growth of complexity. It is not. Indeed, we may say that, if such an implication *were* present, we should probably have to regard our solution as defective, for it would entirely beg the question of how such a relatively complex entity as u_0 (or something fairly close to it) could arise in the first place. Conversely, once we recognise the *possibility* of evolution within the framework of Genetic Relativism, we can at least see how such prior elaboration of the powers of the constructive automata could occur "in principle"; this insight remains valid, at least as a coherent conjecture, even if we have not demonstrated it in operation. This has a possible

advantage in relation to the solution of von Neumann's problem in that it may permit us to work, initially at least, with significantly more primitive constructive automata as the bases of our self-reproducers.

Thirdly, Genetic Absolutism views all the self-reproducers under investigation as connected by a *single* "genetic network" of mutational changes. This is sufficient to solve von Neumann's problem, as stated, which called only for exhibiting the *possibility* of mutational growth of complexity. In practice, however, we are interested in this as a basis for a *Darwinian* growth of complexity. Roughly speaking, this can only occur, if at all, along paths in the genetic network which lead "uphill" in terms of "fitness". If the genetic network is fixed then this *may* impose severe limits on the practical paths of Darwinian evolution (and thus on the practical growth of complexity). Again, once we recognise the *possibility* of evolution within a framework of Genetic Relativism—which offers the possibility, in effect, of changing, or jumping between, *different* genetic networks—the *practical* possibilities for the (Darwinian) growth of complexity are evidently greatly increased.

This last point represents a quite different reason for favouring the framework (or perhaps we may now say "research programme") of Genetic Relativism, and it is independent of the "power" of particular core constructive automata. In particular, even if we can exhibit a single full blown general constructive automaton, which yields a mutationally connected set of self-reproducers spanning (virtually) every possible behavior supported in the system, there could still be advantages, from the point of view of supporting Darwinian evolution, in identifying alternative constructive automata, defining alternative genetic networks (viewed now as evolutionarily accessible pathways through the space of possible automaton behaviors).

Indeed, this need not be all that difficult to do: it provides a particular reason to consider combining a basic constructive automaton with a turing machine (or something of similar computational powers): the latter is arranged so that it "pre-processes" the description tape in some (turing computable) fashion. The program of the turing machine could then effectively encode a space of alternative genetic languages (subject to the primitive constructional abilities of the original constructive automaton); with moderately careful design, it should be possible to open up an essentially infinite set of constructive automata, which are themselves connected under mutation (of the program for the embedded turing machine—another tape of some sort), thus permitting a multitude of *different* genetic networks for potential exploitation by a Darwinian evolutionary process. This should greatly enhance the possibilities for Darwinian evolution of *any* sort, and thus, in turn, for evolution involving the growth of complexity.

This particular idea seems to have been anticipated by Codd:

> A further special case of interest is that in which both a universal computer and a universal constructor [*sic*] exist and the set of all tapes required by the universal constructor is included in the Turing domain T. For in this case it is possible to present in coded form the specifications of configurations to be constructed and have the universal computer decode these specifications ... Then the universal constructor can implement the decoded specifications. Codd [6, pp. 13–14]

While Codd did not elaborate on *why* such flexibility in "coding" should be of any special interest, it seems plausible that he had in mind precisely the possibility of opening up alternative genetic networks.

A final consideration here is the "compositionality" of the genetic mapping or language. When tackling von Neumann's problem within the framework of Genetic Absolutism, it was *necessary* to assume a degree of compositionality in the genetic language, to assure that there would exist a range of mutations *not* affecting the core constructive automaton in a self-reproducer; without this assumption it would be difficult, if not impossible, to argue that the set of self-reproducers anchored on this single core general constructive automaton would be connected under mutation. This compositionality assumption is more or less equivalent to the biological hypothesis of *Genetic Atomism*, which holds that genomes may be systematically decomposed into distinct *genes* which, individually, have absolute effects on phenotypic characteristics (see McMullin [13, p. 11], Dawkins [7, p. 271]). This again represents a divergence between von Neumann's pragmatically convenient solution schema for his particular problem, and the realities of the biological world (where any simple Genetic Atomism is quite untenable). I conjecture therefore that, should we wish to move away from a strict Genetic Absolutism in our formal or artificial systems we might well find it useful, if not essential, to abandon simple compositionality in our genetic language(s) (i.e., Genetic Atomism) also. This, in turn, would ultimately lead away from self-reproducer architectures in which there is any simple or neat division between the core constructive automaton and the rest of the automaton (though there might still be a fairly strict separation of the description tape from the rest of the machine—i.e., a genotype/phenotype division).

6 Conclusion: Looking Forwards

By re-examining von Neumann's work in the light of his own description of the problem he was working on, I have tried to show that there is much more substance to it than has been generally recognised. However, as Popper [16] has emphasised, the scientific enterprise is intrinsically iterative: the solution of any given problem gives rise to a new problem situation, which is to say new problems to be addressed. It seems to me that von Neumann's work is particularly fruitful in this regard, as it poses a number of new and profound questions which need to be addressed by the (still) fledgling field of Artificial Life. Among these are:

1. Can we clarify (or even formalise) the notion of "biological" or "adaptive" complexity. This of course is not specifically a problem for Artificial Life, but rather underlies the entire discipline of evolutionary biology [11].

2. What is the precise significance of the self-reproductive capability of von Neumann's machines? Technically, based on the idea of a "general constructive automaton", growth of complexity *per se* could take place in isolation from self-reproduction. One simple scenario here would be to imagine u_0 being made to simply grind through the (countable) infinity of all description tapes, constructing every described automaton in turn. This would require only a trivial extension of the capabilities of u_0. While this would fail in practice due to the essential fragility of von Neumann's automata (discussed further below) it is not clear whether there is any *fundamental* problem with this general idea. Nonetheless, it seems clear that if the growth of complexity is to involve *Darwinian* evolution, then self-reproduction surely is a necessary additional requirement.

3. The problem of identity or individuality. In the alife systems discussion above, the question of what constitutes a distinct individual "machine" or "automaton" is addressed in a completely ad hoc manner. In a real sense, these putative individuals exist as such only in the eye of the human observer. Whereas, the notion of a self-defining identity, which demarcates itself from its ambiance, is arguably essential to the very idea of a biological organism. Some early and provocative work on this question, including simulation models overtly reminiscent of von Neumann's CA, was carried out by Varela et al. [21]. However, there has been relatively little further development of this line since.

4. The evolutionary boot-strapping problem: in the von Neumann framework, at least, u_0 is already a very complicated entity. It certainly *seems* implausible that it could occur spontaneously or by chance. Similarly, in real biology, the modern (self-consistent!) genetic system could not have plausibly arisen by chance [5]. It seems that we must therefore assume that something like u_0 (or a full blown genetic system) must itself be the product of an extended evolutionary process. Of course, the problem with this—and a major part of von Neumann's own result—is that it seems that something like a genetic system is a *pre-requisite* to any such evolutionary process. As noted, a framework of Genetic Relativism, which envisages evolutionary modification of the genetic machinery itself, *may* contribute to resolving this problem.

5. Perhaps most importantly of all, what further conditions are required to enable an *actual*, as opposed to merely *potential* growth of complexity? It was well known even to von Neumann himself that his system would not *in practice* exhibit any evolutionary growth of complexity. The proximate reason is that, in his CA framework, all automata of any significant scale are extremely *fragile*: that is, they are very easily disrupted even by minimal perturbation from the external environment. The upshot is after the completion of even a single cycle of self-reproduction the parent and offspring would almost immediately perturb, and thus effectively destroy, each other. This can be avoided by ad hoc mechanisms to prevent all interaction (again, this was suggested by von Neumann, and since shown in practice by Langton and others). However, eliminating interaction eliminates the grist from the darwinian mill, and is thus a non-solution to the substantive problem (growth of complexity). `Tierra` offers a somewhat different approach, whereby the integrity of the automata (process images) is "protected" by the underlying operating system, while still allowing some limited, but significant, forms of interaction. This has been very fruitful in allowing the demonstration of *some* significant evolutionary phenomena. However, any serious claim to substantively model real biological organisms will inevitably have to confront their capacity for *self* maintenance and repair in the face of continuous perturbation and material exchange with their environments. This, in turn, is clearly also related to the earlier problem of biological individuality.

In conclusion, it seems to me that von Neumann's work in Artificial Life—properly understood—is as profound and important today as it was half a century ago; and that it should continue to provide structure, insight and inspiration to the field for many years to come.

Acknowledgements

Many of the ideas presented here were first explored in my PhD thesis [12]; I was privileged to carry out that work under the supervision of the late John Kelly. A large number of people have since helped and challenged me in trying to refine the analysis and arguments. I am particularly indebted to Chris Langton, Glen Ropella, Noel Murphy, Tim Taylor, Mark Bedau and Moshe Sipper; the latter also provided critical encouragement to finally write this particular paper. I am grateful also to the reviewers for comprehensive and constructive criticism. Financial support for the work has been provided by the Research Institute in Networks and Communications Engineering (RINCE) at Dublin City University.

References

[1] E. R. Berlekamp, J. H. Conway, and R. K. Guy. What is life? In *Winning Ways for your Mathematical Plays*, volume 2, chapter 25, pages 817–850. Academic Press, London, 1982.

[2] Arthur W. Burks, editor. *Theory of Self-Reproducing Automata [by] John von Neumann*. University of Illinois Press, Urbana, 1966.

[3] Arthur W. Burks, editor. *Essays on Cellular Automata*. University of Illinois Press, Urbana, 1970.

[4] Arthur W. Burks. Von neumann's self-reproducing automata. In *Essays on Cellular Automata* Burks [3], pages 3–64 (Essay One).

[5] A. G. Cairns-Smith. *Genetic Takeover and the Mineral Origins of Life*. Cambridge University Press, Cambridge, 1982.

[6] E. F. Codd. *Cellular Automata*. ACM Monograph Series. Academic Press, Inc., New York, 1968.

[7] Richard Dawkins. *The Selfish Gene*. Oxford University Press, Oxford, new edition, 1989.

[8] Gabor T. Herman. On universal computer-constructors. *Information Processing Letters*, 2:61–64, 1973.

[9] Christopher G. Langton. Self-reproduction in cellular automata. *Physica*, 10D:135–144, 1984.

[10] Christopher G. Langton. Artificial life. In Christopher G. Langton, editor, *Artifical Life*, volume VI of *Series: Sante Fe Institute Studies in the Sciences of Complexity*, pages 1–47. Addison-Wesley Publishing Company, Inc., Redwood City, California, 1989.

[11] John Maynard Smith. The status of neo-Darwinism. In C. H. Waddington, editor, *Towards a Theoretical Biology, 2: Sketches*, pages 82–89. Edinburgh University Press, Edinburgh, 1969. This paper is also accompanied by various addenda and comments (pages 90–105 of the same work).

[12] Barry McMullin. *Artificial Knowledge: An Evolutionary Approach*. PhD thesis, Ollscoil na hÉireann, The National University of Ireland, University College Dublin, Department of Computer Science, 1992.
`http://www.eeng.dcu.ie/~alife/bmcm_phd/`

[13] Barry McMullin. Essays on darwinism. 3: Genic and organismic selection. Technical Report `bmcm9203`, School of Electronic Engineering, Dublin City University, Dublin 9, Ireland, 1992.
`http://www.eeng.dcu.ie/~alife/bmcm9203/`

[14] Marvin L. Minsky. *Computation: Finite and Infinite Machines*. Prentice-Hall Series in Automatic Computation. Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1967.

[15] Umberto Pesavento. An implementation of von Neumann's self-reproducing machine. *Artificial Life*, 2(4):337–354, 1995.

[16] Karl R. Popper. *Unended Quest*. Fontana/William Collins Sons & Co. Ltd, Glasgow, 1976.

[17] Thomas S. Ray. An approach to the synthesis of life. In Christopher G. Langton, Charles Taylor, J. Doyne Farmer, and Steen Rasmussen, editors, *Artifical Life II*, volume X of *Series: Sante Fe Institute Studies in the Sciences of Complexity*, pages 371–408. Addison-Wesley Publishing Company, Inc., Redwood City, California, 1992. Proceedings of the workshop on Artificial Life held February, 1990, in Sante Fe, New Mexico.

[18] A. H. Taub, editor. *John von Neumann: Collected Works. Volume V: Design of Computers, Theory of Automata and Numerical Analysis*. Pergamon Press, Oxford, 1961.

[19] Timothy John Taylor. *From Artificial Evolution to Artificial Life*. PhD thesis, University of Edinburgh, 1999.
`http://www.dai.ed.ac.uk/daidb/homes/ timt/papers/thesis/html/main.html`

[20] J. W. Thatcher. Universality in the von neumann cellular model. In Burks [3], pages 132–186 (Essay Five).

[21] Francisco J. Varela, Humberto R. Maturana, and R. Uribe. Autopoiesis: The organization of living systems, its characterization and a model. *BioSystems*, 5:187–196, 1974.

[22] John von Neumann. First draft of a report on the EDVAC. The (corrected) version at the URL below has been formally published in the IEEE Annals of the History of Computing, **15**(4), 1993, 1945.
`ftp://isl.stanford.edu/pub/godfrey/ reports/vonNeumann/vnedvac.pdf`

[23] John von Neumann. Theory and organization of complicated automata. In Burks [2], pages 29–87 (Part One). Based on transcripts of lectures delivered at the University of Illinois, in December 1949. Edited for publication by A.W. Burks.

[24] John von Neumann. The general and logical theory of automata. In Taub [18], chapter 9, pages 288–328. Delivered at the Hixon Symposium, September 1948; first published 1951 as *pages 1–41 of:* L. Jeffress, A. (ed), *Cerebral Mechanisms in Behavior*, New York: John Wiley.

Author Contact Information

Barry McMullin

DCU Alife Laboratory
Research Institute for Networks and Communications Engineering (RINCE)
Dublin City University
Dublin 9
Ireland.

Voice: +353-1-700-5432
Fax: +353-1-700-5508
E-mail: Barry.McMullin@dcu.ie
Web: http://www.eeng.dcu.ie/~mcmullin

Online Retrieval

The resources comprising this paper are retrievable in various formats via:

- http://www.eeng.dcu.ie/~alife /bmcm-alj-2000/

The final version of this article has been published in Artificial Life, Vol. 6, Issue 4, Fall 2000, pp. 347–361. Artificial Life is published by The MIT Press.

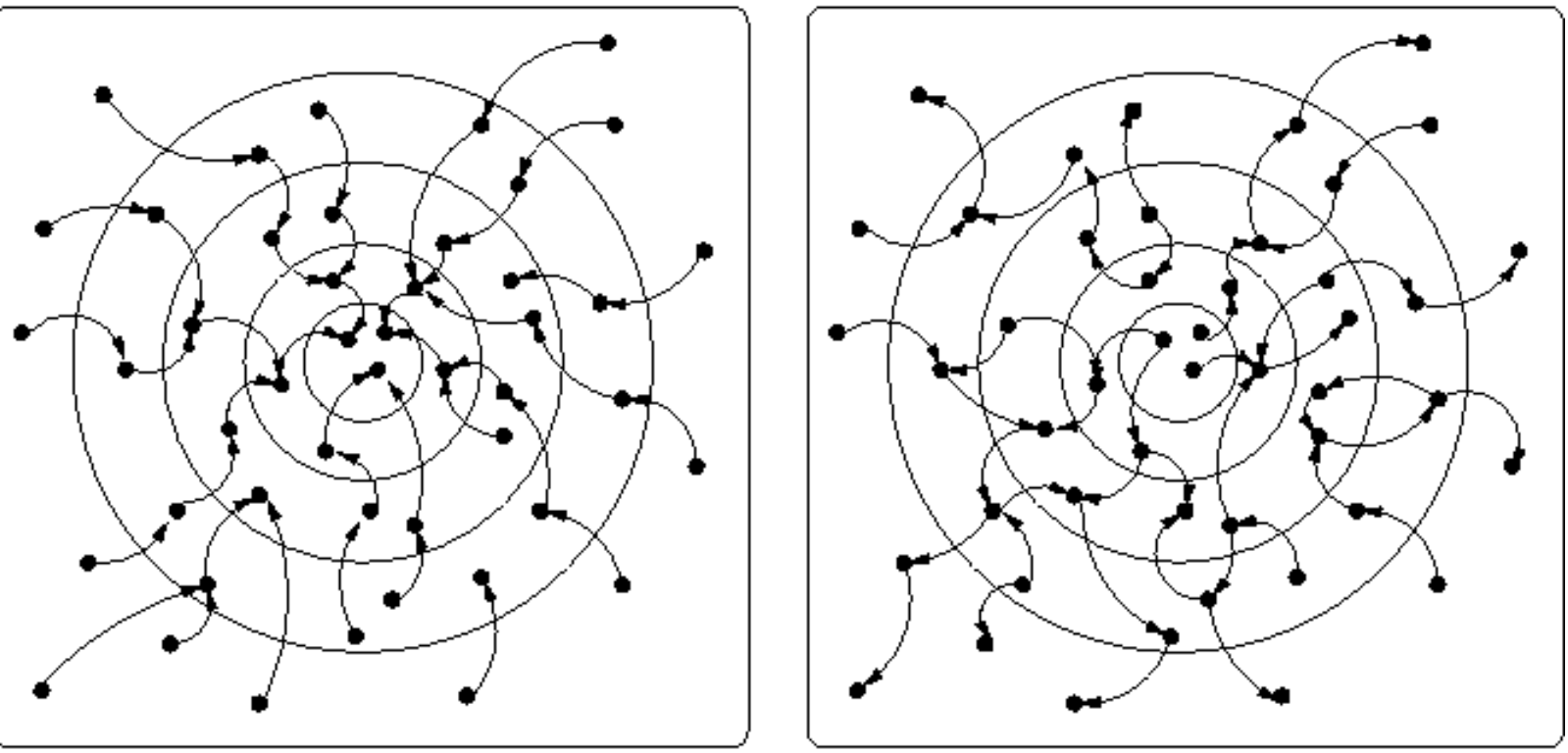

Figure 1: Evolutionary growth of complexity.