

Who Needs Genomes?

Barry McMullin
Dublin City University
Barry.McMullin@rince.ie

Tim Taylor
University of Abertay Dundee
Tim.Taylor@abertay.ac.uk

Axel von Kamp
Dublin City University
Axel.vonKamp@rince.ie

Abstract

The first detailed mechanistic models for genome based reproduction were developed by John von Neumann in the period 1948-1953 (von Neumann, 1948, 1949; Burks, 1966). While these models were extremely abstract, subsequent elaboration of the structure and function of DNA proved von Neumann's designs to have been strikingly prescient. However, some significant questions still remain as to the specific benefits of this particular reproductive architecture. These questions are relevant both to understanding the evolutionary emergence of such systems, and their proper role in engineered or synthetic evolutionary systems. This paper will review these issues, and present some preliminary results of novel evolutionary experiments in the *Tierra* system (Ray, 1992), where artificial "organisms" are deliberately engineered to have an *evolvable* genetic architecture.

The Problem Situation

This paper is concerned with evolutionary systems and their evolvability. By "evolutionary system" we mean a system satisfying the abstract conditions for Darwinian evolution: reproduction with heritable variation, in a finite world, giving rise to natural selection. "Evolvability" is a more nebulous term; for our purposes, it roughly connotes the distinction between evolutionary systems which sustain a spontaneous and apparently open-ended growth of complexity—and those which do not. We have, by now, many examples of the latter, but only one of the former. We suspect this to be rather a deep problem; but we will attempt to scratch its surface by probing the relationship between certain distinguishable modes of reproduction and consequent evolutionary potential.

Reproduction

Reproduction with heritable variation is at the heart of any Darwinian evolutionary process. Through the inevitable collision between Malthusian growth and limited resources, it establishes the conditions

for natural selection—quasi-deterministic displacements of one lineage by another. But more importantly for our purposes here, the patterns of variation establish the potential for continuing innovation, and ultimately, continuing growth of complexity.

Maynard Smith and Szathmáry (1999, pp. 7–9) make a useful distinction between "limited" and "unlimited" heredity. The former exhibits heritable variation, but the total number of distinct reproductive variants is finite and rather small; the latter encompasses an indefinitely large number of distinct reproductive variants. We will be concerned throughout with evolutionary systems having unlimited heredity.

We know of just two clearly distinguished modes or processes for achieving reproduction with (unlimited) heritable variation. Many different terms have been coined for these; we will use "template" and "genetic" reproduction. Genetic reproduction is the more complicated—not least in the sense that it relies on already having a subsidiary template reproduction process at its disposal. Genetic reproduction also seems to be the more "powerful", in several distinct ways. Our primary purpose in this paper is to distinguish and elucidate these as clearly as possible.

Templates

By template reproduction¹ we mean a process whereby an offspring is constructed by *copying* a parent: i.e., the parent serves precisely as a "template". The reproductive entity is conceived as some sort of composite system, consisting of a particular configuration or arrangement of components. It must be presumed, of course, that the requisite "raw materials"—the components to be assembled into the offspring—are available; and the process may or may not rely on some extrinsic "machinery" (catalyst) which is not itself reproduced. It is precisely because template reproduction *is* a copying process that it supports heritable variation: any variation in

¹Also often called simply "replication".

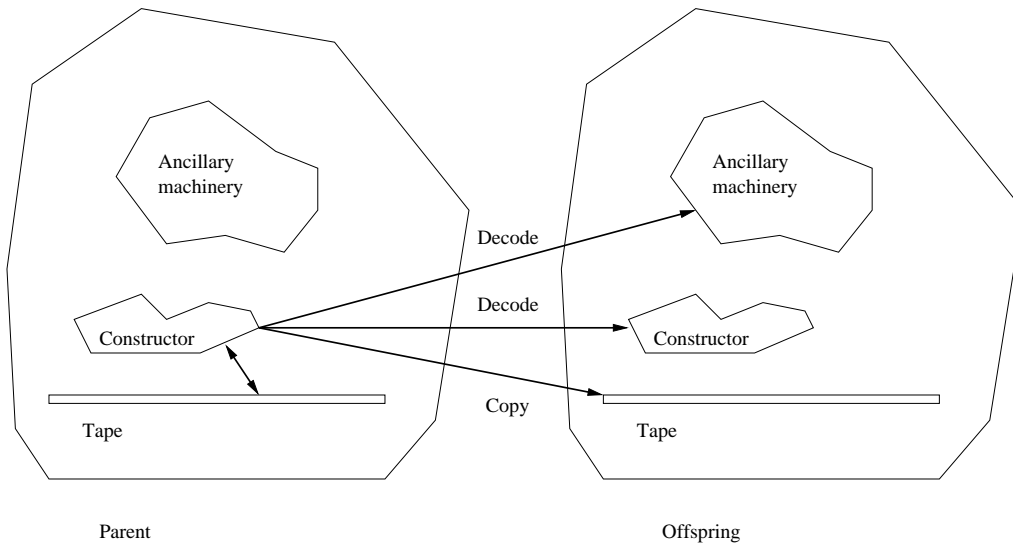


Figure 1: Von Neumann (Genetic) Reproducer

a parent, howsoever caused, which is recognized and preserved by the particular copying process, will be inherited. For unlimited heredity we require that the class of variant entities that can be faithfully reproduced in this way is indefinitely large. Of course this still allows that there may be many forms of variation that are *not* heritable (indeed that may cause reproduction to fail altogether).

The practical realisation of template reproduction requires a process that can recognize the individual components of the parent, and their interconnection or configuration; and can then cause matching components to be assembled in a matching configuration. For unlimited heredity, this process must be able to work over an indefinitely large set of distinct parent entities.

Abstractly, we can suppose the set of distinct reproductive variants to be enumerated—i.e., each tagged with an integer identifier. The “normal” offspring of any given parent would then be identified with the same tag; variant or mutant offspring would be identified with a different tag. Such numeric tags can be used to apply information theoretic ideas to template reproduction; in particular, one can view the reproductive entity as *carrying* information, and as *communicating* or *transmitting* information to its offspring.

Natural examples of template reproduction with unlimited heredity would be RNA and DNA replication. Some familiar artificial examples would be photocopying, and copying of configurations of various magnetic and optical media (cassette and VCR tapes, computer diskettes, CDs, DVDs etc.).

Genomes

Genetic reproduction, in the abstract form originally proposed by von Neumann (1948), is illustrated in Figure 1. The Von Neumann reproducer is composed of three major subsystems: the *tape* is a more or less quiescent information carrier, subject to template style reproduction; the *constructor* is a machine capable both of copying a tape (to yield another with the same information content) and of “decoding” a tape (to yield more or less arbitrary machinery, as described by the information on the tape);² and finally the *ancillary machinery* is just a name for all additional machinery constituting the reproducer (i.e., with functionality not directly related to reproduction). In biological parlance, the tape can be regarded as the genome of the reproducing entity, while the constructor and ancillary machinery together constitute the phenotype. The reproductive cycle is driven by the constructor. The process is as follows:

1. The tape is copied.
2. A section of tape is decoded to yield a new constructor subsystem; a separate section of tape is decoded to yield new ancillary machinery. For simplicity, these are constrained to be initially quiescent.

²This is as von Neumann formulated his design, but it may be noted that nothing important hangs on the constructor having this double functionality; the tape copying could be mediated by completely separate machinery, or even no machinery at all, without materially affecting any of the analysis which follows.

3. The new tape, constructor and ancillary machinery are assembled together, and activated.

It is crucial to note that while the offspring tape is identical (in the informational sense) to the parental tape precisely because it is copied from it, the relationship between the parental and offspring phenotypes (constructor plus ancillary machinery) is *not* based on a copying process, and is in fact rather subtle. The offspring phenotype results from a *decoding* of the tape, mediated by the parental constructor. So the offspring phenotype will be similar or identical to the parental phenotype *only* if the parental tape “happens” to carry an accurate *encoding* of this parental phenotype.³ In practice, in order to design a reproducing machine with this architecture, von Neumann *first* designed the phenotype in detail, and then deliberately contrived a suitable tape by manually *encoding* that phenotype. With this in place, the whole system can then successfully reproduce.

In any case, having once devised such a reproducer, which “breeds true”, it is clear that certain kinds of variants, or mutants, will also be capable of breeding true—i.e., that this architecture allows for a *new* form of unlimited, heritable, variation, at a level over and above that supported by the underlying *template* reproduction of the tapes. Indeed, achieving this was von Neumann’s primary motivation (von Neumann, 1949, Fifth lecture, p. 86).

Specifically, such variants can arise through rather arbitrary perturbations of that part of the parental tape which encodes for the ancillary machinery. Provided that this happens before the reproductive cycle starts, the result will be an offspring with a variant genome with a *matching* variant phenotype, such that this combined variation will indeed recur, or breed true, in this new lineage.

The close parallels between this abstract model and our modern molecular level understanding of real biological reproduction, including the “central dogma” of one directional information flow from genome to phenotype, should be clear. This seems at least somewhat remarkable considering that von Neumann’s model was first formulated in 1948, some five years before even the double helix structure of DNA was identified by Watson and Crick (1953).

So Why Bother?

Nonetheless, the bare fact that von Neumann’s model proved to be remarkably prescient does not in itself explain *why*, or in what circumstances, this

³Of course, this encoding must be relative to the particular decoding implemented by the parental constructor. We will return to this point shortly.

genetic mode of reproduction will be useful or appropriate. Von Neumann himself proposed two reasons:

- Copying relies on quiescence or stability of that which is being copied. Or to put it the other way around, if the target structure is itself dynamically changing, then it can change significantly during the “copying” process; on a macroscopic scale the supposed “copy” must then have an indeterminate relationship with the original, despite being the outcome of even a “perfect” copying process (on the microscopic scale).
- Copying, by definition, relies on being able to inspect the components, and their relationships, of whatever structure is to be copied. If the structure is one dimensional—a chain of some sort—then this inspection can be done, in principle at least, without having to actually dismantle the structure. This may even be possible for some two dimensional structures (by inspection in the third dimension—see, for example, Cairns-Smith 1982). However, it becomes effectively impossible for three dimensional structures in a three dimensional space.

Both of these are, of course, directly applicable to natural, biological, reproduction, and they may well be adequate to explain the significance of genetic architecture in that case. However, the situation is different when considering the engineering of reproduction to support *artificial* evolution.

Even in von Neumann’s own most detailed design of such a system (in his 29-state cellular automaton space, von Neumann 1953), the first reason above does not directly apply, because this space allows for essentially arbitrary configurations to be rendered into a “quasi-quiescent” state. The second reason—the dimensionality constraint—*does* apply to this system; it is effectively two dimensional and thus quite strictly limits inspection without disassembly to one dimensional structures. However, this is a self imposed restriction. In fact, if the entities of interest are embedded in some artificially engineered space then the “dimensionality” of interactions can be specified quite arbitrarily.

This can be seen very clearly in a number of recent model evolutionary systems, perhaps the best known of which is Tom Ray’s *Tierra* (Ray, 1992). In this system, the evolutionary actors are small computer programs, inhabiting a shared random access memory or RAM. These are not constrained by either of the two problems identified above. The individual memory locations are quasi-quiescent in a similar sense to von Neumann’s cellular model; and by the very nature of “random access” memory, any

location can be inspected without disturbance and without constraint—the dimensionality of interaction is effectively unlimited.⁴ Because of this, arbitrary program entities in *Tierra*—from the simplest to the most complex—can be made to reproduce simply by copying of the program image in RAM—i.e., template based reproduction. That being the case, there is certainly no obvious reason to invoke the more involved mechanism of genetic reproduction; and, indeed, in the experiments reported by Ray and his co-workers, only template style reproduction has been used.

However, notwithstanding this, we now wish to draw attention to one further distinction between template and genetic style reproduction—which may still mean that the latter is preferable *even* in artificial systems where von Neumann’s original reasons for proposing it need not apply.

Mutatis Mutandis...

Evolvability (in our sense of the possible evolutionary growth of complexity) certainly requires the bare existence of an indefinitely large set of potential reproducers, interconnected by a mutational network (“unlimited heredity”). This is enough to assure that there will be potential mutational trajectories from simpler to more complex entities. This can, of course, be achieved by genetic reproduction; but as illustrated by *Tierra*, it *can* be achieved even by template reproduction.

However, which mutational trajectories will actually be followed will depend critically on the detailed interconnections in this network, and the associated patterns of Darwinian selective displacement; should these result in relatively simple entities being selectively favored over their more or less *immediately* accessible mutational neighbors, then longer term evolutionary potential will be effectively blocked.

Now, in *this* respect there seems to be a significant difference between template and genetic reproduction. In a system relying on template reproduction, there is a fixed, essentially isomorphic, relationship between genotype and phenotype. Accordingly, the mutational connectivity of particular phenotypes is identical with the mutational connectivity of the genotypes—with whatever limitations may result on long term evolutionary dynamics. But in a

⁴Admittedly, this ease of interaction actually makes the problem of realising “quasi-quiescence” rather harder—because any given program entity can effectively disrupt any other. This seriously constrained evolutionary phenomena in earlier systems of this sort, such as the α -universes (Holland, 1976; McMullin, 1992) or Coreworld (Rasmussen et al., 1990). Perhaps the most important innovation in the development of *Tierra* was the introduction of “memory protection” which allowed for control of such interactions.

system based on genetic reproduction, there is a decoupling between genotype and phenotype. There is, of course, a relationship, or mapping, between genotype and phenotype, and this still means that the connectivity of the genotype space implies connectivity of the phenotype space—*unless the mapping between genotypes and phenotypes can itself evolve*. But if this mapping *is* evolvable, then, without any change to the underlying template copying process, or the corresponding connectivity of the genotype space, the connectivity in the phenotype space can change. Since it is phenotypes that give rise to Darwinian selection, this means that the potential for indefinitely long term evolution will now not necessarily be constrained by the fixed connectivity of the genotype space; and thus genetic reproduction—if it allows for variation in the genotype-phenotype mapping—*might*, in principle, give rise to richer evolutionary potential, or evolvability, than any template style system.

What does it mean for the genotype-phenotype mapping to be evolvable? In terms of our earlier schematic diagram of genetic reproduction (Figure 1) the key issue is whether the constructor system is, itself, subject to (unlimited) heritable variation, for it is the constructor that implements the “decoding” or mapping from genotype to phenotype.

This is a more subtle question than may at first appear. It is easy enough to arrange for the constructor subsystem to be constructed by virtue of decoding some particular section of the tape. Accordingly, an alternation or mutation in that section will indeed result in a variant constructor in the offspring. This offspring will thus have both a variant constructor and a corresponding variant genotype, so it would seem that the variation can now breed true, as usual: except that this “correspondence” is as defined by the *parental* genotype-phenotype mapping—and, by stipulation, the *offspring* no longer shares this mapping.

We should note here that von Neumann himself seems to have discounted this possibility completely. He stated explicitly that mutations affecting that part of a descriptor coding for the constructor would result in the production of “sterile” offspring (von Neumann, 1949, p. 86). Clearly, on this specific point, we disagree with von Neumann. We do accept that such mutations might “typically” result in sterile offspring; but we suggest that, in principle at least, they *may* sometimes result in viable offspring, thus initiating lineages with distinctly different evolutionary potential, precisely because of the altered genotype to phenotype mapping.

A Model

We outline here a very preliminary result from exploring this issue in the *Tierra* system. The system was seeded with an ancestor program, designed to reproduce genetically rather than by template copying. The genotype to phenotype mapping consisted of a simple recoding of each allowed machine word by a different word, via a lookup table. The lookup table itself was explicitly coded for in the genotype. Thus, mutations in the section of the genotype coding for the lookup table would result in a different table in the offspring, and thus a different mapping from genotype to phenotype in successive offspring in such a lineage.

A number of experiments have been performed on the subsequent evolutionary behaviour in this system. These will not be presented in detail here: however, we have indeed detected the emergence of new programs with mutated genotype-phenotype mappings (in the sense of mutated translation tables) which, nonetheless, subsequently breed true. To test the degree of change in mapping, we have artificially transplanted the genotype from such a (remote) descendant back into the original ancestor phenotype and verified that it cannot recreate the descendant (precisely because this genotype does not represent the descendant phenotype *relative to the mapping implemented by the ancestor*).

Now the nature of the very simple genotype to phenotype mappings used in these particular experiments means that we do not expect that the changes in mapping reported above would actually amount to evolutionarily significant changes in the mutational connectivity of the phenotype space; but they do concretely demonstrate that evolution in the genotype-phenotype mapping is, at the very least, possible.

Conclusion

We have suggested that while there are some clear and straightforward benefits of genetic reproduction, there may also be additional, and more subtle benefits. Among these may be the extra evolutionary potential that might be opened up if the genotype to phenotype mapping is itself evolvable. This is particularly relevant to the design of artificial evolutionary systems (where the other advantages of genetic reproduction need not apply); but it may also have some significance for natural, biological, evolution, not least in understanding the evolutionary emergence of genetic reproduction in the first place.

References

- Burks, A. W., ed. (1966), *Theory of Self-Reproducing Automata [by] John von Neumann*, University of Illinois Press, Urbana.
- Cairns-Smith, A. G. (1982), *Genetic Takeover and the Mineral Origins of Life*, Cambridge University Press, Cambridge.
- Holland, J. H. (1976), Studies of the Spontaneous Emergence of Self-Replicating Systems Using Cellular Automata and Formal Grammars, in A. Lindenmayer and G. Rozenberg, eds, 'Automata, Languages, Development', North-Holland, New York, pp. 385–404.
- Jeffress, L. A., ed. (1951), *Cerebral Mechanisms in Behavior*, John Wiley, New York.
- Maynard Smith, J. and Szathmary, E. (1999), *The Origins of Life*, Oxford University Press, Oxford.
- McMullin, B. (1992), The Holland α -Universes Revisited, in F. J. Varela and P. Bourguin, eds, 'Proceedings of the First European Conference on Artificial Life', MIT Press, Cambridge, pp. 317–326.
- Rasmussen, S., Knudsen, C., Feldberg, R. and Hindsholm, M. (1990), 'The Coreworld: Emergence and Evolution of Cooperative Structures in a Computational Chemistry', *Physica* **42D**, 111–134.
- Ray, T. S. (1992), An approach to the synthesis of life, in C. G. Langton, C. Taylor, J. D. Farmer and S. Rasmussen, eds, 'Artificial Life II', Addison-Wesley Publishing Company, Inc., Redwood City, California, pp. 371–408.
- Taub, A. H., ed. (1961), *John von Neumann: Collected Works. Volume V: Design of Computers, Theory of Automata and Numerical Analysis*, Pergamon Press, Oxford.
- von Neumann, J. (1948), The General and Logical Theory of Automata, in Taub (1961), chapter 9, pp. 288–328. Delivered at the Hixon Symposium, September 1948; first published as pp. 1–41 of Jeffress (1951).
- von Neumann, J. (1949), Theory and Organization of Complicated Automata, in Burks (1966), pp. 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.
- von Neumann, J. (1953), The Theory of Automata: Construction, Reproduction, Homogeneity, in Burks (1966), pp. 89–250. Based on an unfinished manuscript by von Neumann. Edited for publication by A.W. Burks.
- Watson, J. and Crick, F. (1953), 'A Structure for Deoxyribose Nucleic Acid', *Nature* **171**, 737–738.