

Creativity in Evolution: Individuals, Interactions and Environments

Tim Taylor

Institute of Perception, Action and Behaviour
Division of Informatics, University of Edinburgh
tim.taylor@ed.ac.uk

Abstract

This chapter addresses the nature of open-ended evolutionary processes, and the related, but more subtle, issue of how fundamental novelty (i.e. creativity) can arise in such processes. A number of existing artificial evolutionary systems, such as Tierra (Ray, 1991), are analysed in this context, but it is found that the theoretical grounding upon which they are based does not usually consider all of the relevant issues for creative evolution. The importance of considering the design of the environment, and of interactions between individuals, as well as the design of the individuals themselves, is emphasised. The properties of a hypothetical 'proto-DNA' structure—a suitable seed for an open-ended, and creative, evolutionary process—are discussed. A number of open questions relating to these issues are highlighted as useful areas of future research. Finally, a paradigm for an evolutionary process described by Waddington (1969) is described. It is suggested that this might represent a suitable starting place for a more unified and productive exploration of these issues using synthetic (artificial life) modelling techniques.

1 Introduction

This chapter addresses the question: What are the basic design considerations for creating an artificial evolutionary system that displays the sort of creativity observed in biological evolution? I am therefore specifically considering evolutionary systems which possess an inherent ability to be creative, rather than those in which creativity is achieved by interactions with a human observer. I start by discussing what I mean by creativity in this context, and how it relates to open-ended evolution. I then discuss various issues concerning the design of artificial evolutionary systems and their capacity for creative evolution. The discussion emphasises that it is necessary to consider not just the design of individuals, but also the sort of environments in which they live, and how individuals can interact with each other and with the physical (i.e. abiotic) environment. Much of this discussion is presented in relation to a hypothetical structure (which I refer to as 'proto-DNA') that would be suitable for acting as a robust initial seed for an open-ended, creative evolutionary process. I go on to discuss how these issues should be integrated into a unifying framework in which the study of creative artificial evolutionary systems can be developed.

2 Creativity and Open-Ended Evolution

Most forms of artificial evolutionary system are designed to be used as optimisation tools; the course of evolution

is guided by an extrinsically defined fitness function that preferentially selects individuals that are deemed to be 'fit' according to some specific criterion. Examples include genetic algorithms (Holland, 1975), genetic programming (Koza, 1992), and similar techniques. In this type of system, the evolving individuals move towards a predefined, and usually static, fitness peak, and when this peak has been reached, they generally stay there.

In contrast, some other evolutionary systems have a less determinate feel. These include models of *co*-evolutionary processes of one form or another, where the success of organisms in one population depends upon the success of organisms in another, coevolving population. Examples of this type of work include systems described by Hillis (1990), Sims (1994), Miller and Cliff (1994), and Floreano et al. (1998). Hillis, for example, coevolved a population of algorithms for sorting lists of numbers, together with a population of lists which were used to test the algorithms. The idea was that the algorithms were rewarded for correctly sorting the test lists, whereas the test lists were rewarded for baffling the algorithms. Therefore as the algorithms evolved to better deal with the test cases, so the test cases evolved to present harder challenges to the algorithms. One population spurred on the other to higher fitness, and the algorithms obtained by this method were indeed consistently better and faster than those obtained using a fixed set of test cases.

However, these coevolutionary studies are geared towards producing organisms which are good at performing a particular task. To this end, the coevolving organisms are still generally competing in some pre-specified

(extrinsically defined) game, and they are not given the potential for developing entirely *new* games to play.

Another group of models has moved even further from the idea of extrinsically defined fitness functions, dispensing altogether with the notion of modelling evolution towards any sort of high-level goal. Examples include models by Barricelli (1957), Conrad and Pattee (1970), Packard (1988), Ray (1991), Adami and Brown (1994) and Holland (1995). In these systems, individuals are competing for one or more limited resources which they require in order to survive and propagate (e.g. memory or CPU-time). The fact that these resources are limited induces natural (intrinsic) selection for those individuals that out-compete their neighbours. These systems have more of an open-ended nature, because the individuals are not evolving towards any predefined high-level goal; they are being selected for their ability to win the limited resources, but this ability is measured *relative to (some or all of) the other individuals in the population*. Hence, an individual's 'fitness' changes as new individuals are born and existing ones die. As the biotic environment of an individual (i.e. the other individuals in the population) changes, that individual (or its offspring) must adapt in order to survive. This adaptation, in turn, causes the environment experienced by other organisms to change, so the population is in a constant state of flux. This scenario is equivalent to Van Valen (1973)'s Red Queen hypothesis for indefinite evolutionary change in biological ecosystems.

For promoting open-ended evolution, the importance of individuals being part of the environment experienced by other individuals has also been emphasised by some members of the artificial life community, e.g. Ray (1991), Arthur (1994) and Bedau (1998). However, the theoretical considerations driving the design of the above systems have focussed almost exclusively on properties of individuals (e.g. the self-reproduction process). Little is said, from a theoretical point of view, of how the environment should be constructed (including how individuals form part of the environment for other individuals), or how individuals should be allowed to interact.

Some of these latter systems can be regarded as modelling 'open-ended evolution', in the sense that new, adaptively successful individuals continuously appear in the populations—evolutionary activity does not peter out.¹ However, the *kinds* of evolutionary innovation observed in these systems are generally fairly restricted. For example, the evolutionary innovations observed in experiments with Tom Ray's Tierra platform (described in more detail in Section 3.2) fall into two broad categories: 'ecological solutions' and 'optimisations' (Ray, 1997), but the limited interactions between individuals in Tierra restricts the range of possible innovations even within these categories. Much has been said of the evolution of parasites² and related ecological phenomena in Tierra (e.g.

¹Although even in these systems it is debatable whether this can continue indefinitely.

²That is, short programs which are unable to reproduce by them-

Ray, 1991), but the fact that they appear is due to some fairly specific aspects of the system's design and of the particular way in which the ancestral self-reproducing programs were written; these phenomena emerged only because it was very easy for evolution to discover them (see, e.g., Taylor, 1999). In short, it is hard to escape the feeling that most of these systems are only capable of producing innovations of the 'more-of-the-same' variety (e.g. more optimised code), rather than anything fundamentally new.

It is hard to be precise about what counts as 'fundamentally new', but I am referring to the ability of individuals to interact with their (biotic and abiotic) environment with few restrictions, and to evolve mechanisms for sensing new aspects of this environment and for interacting with it in new ways. This includes the ability of individuals to utilise new physical modalities (e.g. sound, light, electrical conductance) which they previously did not use, to develop new functional relationships with their environment (e.g. the ability to fly) and also for the very notion of individuality to change in radical ways (e.g. the evolution of multicellular organisms from unicellular ones). It is these sorts of evolutionary innovations which I am labelling 'creative'. Creativity is therefore distinct from open-endedness; a system capable of open-ended evolution is not necessarily creative. Biological evolution has managed all of these feats, so the question is how to instill similar capacities into artificial evolutionary systems.

In the following sections I analyse the design of artificial evolutionary systems (specifically, those with intrinsic selection) with respect to open-ended evolution. I also consider how the capacity for *creative* evolution can be secured. The analysis emphasises the need for the explicit consideration of environments and of interactions as well as of individuals.

3 Design Issues

I begin this section by introducing von Neumann's work on the logic of self-reproduction. Next I discuss Ray's Tierra model in a bit more detail. I then analyse self-reproduction in a number of artificial evolutionary systems in terms of von Neumann's proposed architecture. Finally, in Sections 3.4 to 3.6, I discuss issues relating to phenotypic properties, and the relationship between individuals and the environment in artificial systems.

3.1 Von Neumann's Architecture for Self-Reproduction

In the late 1940s and early 1950s, John von Neumann devoted considerable time to the question of how complicated machines could evolve from simple machines.³

selfes, but do so by reading code from neighbouring programs.

³Von Neumann had difficulties in defining precisely what the term 'complicated' meant. He said "I am not thinking about how involved the

Specifically, he wished to develop a formal description of a system that could support self-reproducing machines which were robust in the sense that they could withstand some types of mutation and pass these mutations on to their offspring. Such machines could therefore participate in a process of evolution.

Inspired by Turing (1936)'s earlier work on universal computing machines, von Neumann devised an architecture which could fulfil these requirements. The machine he envisaged was composed of three subcomponents (von Neumann, 1966):

1. A general *constructive* machine, **A**, which could read a description $\phi(\mathbf{X})$ of another machine, **X**, and build an instance of **X** from this description:

$$\mathbf{A} + \phi(\mathbf{X}) \rightsquigarrow \mathbf{X} \quad (1)$$

(where $+$ indicates a single machine composed of the components to the left and right suitably arranged, and \rightsquigarrow indicates a process of construction.)

2. A general *copying* automaton, **B**, which could copy the instruction tape:

$$\mathbf{B} + \phi(\mathbf{X}) \rightsquigarrow \phi(\mathbf{X}) \quad (2)$$

3. A *control* automaton, **C**, which, when combined with **A** and **B**, would first activate **B**, then **A**, then link **X** to $\phi(\mathbf{X})$ and cut them loose from (**A** + **B** + **C**):

$$\mathbf{A} + \mathbf{B} + \mathbf{C} + \phi(\mathbf{X}) \rightsquigarrow \mathbf{X} + \phi(\mathbf{X}) \quad (3)$$

Now, if we choose **X** to be (**A** + **B** + **C**), then the end result is:

$$\begin{aligned} \mathbf{A} + \mathbf{B} + \mathbf{C} + \phi(\mathbf{A} + \mathbf{B} + \mathbf{C}) &\rightsquigarrow \\ \mathbf{A} + \mathbf{B} + \mathbf{C} + \phi(\mathbf{A} + \mathbf{B} + \mathbf{C}) &\quad (4) \end{aligned}$$

This complete machine plus tape, [**A** + **B** + **C** + $\phi(\mathbf{A} + \mathbf{B} + \mathbf{C})$], is therefore self-reproducing. From the point of view of the evolvability of this architecture, the crucial feature is that we can add the description of an arbitrary additional automaton **D** to the input tape. This gives us:

$$\begin{aligned} \mathbf{A} + \mathbf{B} + \mathbf{C} + \phi(\mathbf{A} + \mathbf{B} + \mathbf{C} + \mathbf{D}) &\rightsquigarrow \\ \mathbf{A} + \mathbf{B} + \mathbf{C} + \mathbf{D} + \phi(\mathbf{A} + \mathbf{B} + \mathbf{C} + \mathbf{D}) &\quad (5) \end{aligned}$$

Furthermore, notice that if the input tape $\phi(\mathbf{A} + \mathbf{B} + \mathbf{C} + \mathbf{D})$ is mutated in such a way that the description of automaton **D** is changed, but that of **A**, **B** and **C** are unaffected—that is, the mutated tape is $\phi(\mathbf{A} + \mathbf{B} + \mathbf{C} + \mathbf{D}')$ —then the result of the construction will be:

$$\begin{aligned} \mathbf{A} + \mathbf{B} + \mathbf{C} + \mathbf{D} + \phi(\mathbf{A} + \mathbf{B} + \mathbf{C} + \mathbf{D}) &\xrightarrow{\text{mutation}} \\ \mathbf{A} + \mathbf{B} + \mathbf{C} + \mathbf{D}' + \phi(\mathbf{A} + \mathbf{B} + \mathbf{C} + \mathbf{D}') &\quad (6) \end{aligned}$$

object is, but how involved its purposive operations are. In this sense, an object is of the highest degree of complexity if it can do very difficult and involved things." von Neumann (1966).

The reproductive capability of the architecture is therefore robust to some mutations (specifically, those mutations which only affect the description of **D**), so the machines are able to evolve. Von Neumann pointed out that the action of the general copying automaton, **B**, was the decisive step which gave his architecture the capacity for evolving machines of increased complexity, because **B** is able to copy the description of any machine, no matter how complicated (von Neumann, 1966, p.121). This ability is clearly demonstrated in Reaction 5 above.

The original implementation envisaged by von Neumann was a constructive system, which Burks has referred to both as the 'robot model' and as the 'kinematic model' (Aspray and Burks, 1987, p.374). However, von Neumann decided that the system was too complicated to capture in a set of rules that were both simple and enlightening, so he turned his attention to developing the cellular automata (CA) framework with Stanislaw Ulam. Von Neumann described the detailed design of a self-reproducing machine in a cellular automata space, according to the architecture described above.⁴ In this CA model, each of the basic components of von Neumann's architecture, **A**, **B**, **C** and ϕ , were represented as particular configurations of cell states within a two-dimensional lattice of cells, and the action of the cells was defined (as in all CA models) by the particular transition functions used to determine how a cell's state changed over time. Recently, a slightly modified and simplified version of this design was successfully implemented on a computer (Pesavento, 1995). One of the major achievements of von Neumann's work was to clarify the logical relationship between *description* (the instruction tape, or genotype), and *construction* (the execution of the instructions to eventually build a new individual, or phenotype) in self-replicating systems. However, as already mentioned and as emphasised recently by McMullin (1992), his work was always within the context of self-replicating systems which would also possess great *evolutionary* potential.

3.2 Tierra

An artificial evolutionary system of a somewhat different design to von Neumann's that has received a great deal of attention in the last decade is Tom Ray's Tierra model (Ray, 1991)—mentioned in Section 2. Tierra is an implementation of a virtual parallel computer that can simulate the concurrent execution of many hundreds of programs. The programs are written in a specially designed language that is both robust and simple. Programs written

⁴The general constructive machine **A** of this design is often referred to as a 'universal constructor'. However, this term should be used with caution; from the above description of the architecture it is clear that **A** can build any machine **X** that can be described upon a tape $\phi(\mathbf{X})$. For cellular automaton models it can be proved that there are some configurations that the universal constructor cannot build (e.g. Moore, 1962; Myhill, 1963). These are referred to as 'Garden of Eden' configurations, as the only way they may exist is if they are programmed in as the initial state of the space at time zero.

in this language can be mutated (i.e. random changes can be made to them) without causing the computer to crash.

An evolutionary run commences with the introduction of an *ancestor* program into the otherwise empty memory. The ancestor is a hand-written self-replicator which produces another copy of itself in the computer's memory when it is run. At each iteration of the system, each program in the computer's memory is allowed to execute a certain number of instructions. A small element of stochastic behaviour is associated with the execution of the machine instructions, e.g. an `add` instruction which usually adds one to its operand may occasionally add zero or two instead, or a `copy` instruction may sometimes mutate a byte of the data it is copying. The programs are also subject to point random mutations at a given low rate. In either of these ways, as a run proceeds variations of the ancestor program begin to appear. If a variation retains the ability to produce a copy of itself, then it too may be retained in the population of programs over a number of generations. As the available memory begins to fill, a 'reaper' operation is performed to kill off a number of the programs. Programs which perform operations which cause their flag to be set⁵ are killed off quicker than others (by being advanced up the 'reaper queue'), but otherwise the order in which programs are killed off is largely determined by their age.

As mentioned in Section 2, a number of interesting results have been obtained from such evolutionary runs. For example, 'parasites' have appeared—short pieces of code which run another program's copying procedure in order to copy themselves. Hyper-parasites (parasites of parasites) have also been observed, along with a number of other interesting ecological phenomena (Ray, 1991).

Although *Tierra* was designed to study evolution, and in particular (originally, at least) the evolution of multicellular organisms from unicellular ones, it was not built around any particular theory of what the important features of this transition might have been. There are therefore no coherent theoretical reasons for deciding which features should be modelled, and which should be left out. This weakness is not specific to *Tierra*, but is shared by most, if not all, of the other *Tierra*-like systems which have emerged over the last decade (e.g. Adami and Brown, 1994; Taylor, 1997).

In describing the philosophy behind the *Tierra* system, Ray explains that:

“... rather than attempting to create prebiotic conditions from which life may emerge, this approach involves engineering over the early history of life to design complex evolvable organisms, and then attempting to create conditions that will set off a spontaneous evolutionary process of increasing diversity and complexity of organisms” (Ray, 1991, p.373).

⁵Examples include issuing a `jmp` instruction with a template pattern for which no match can be found, and attempting to write to a memory address for which the program does not have write access.

However, in order to 'engineer over' several billion years of evolution, we would need to have a very good idea of the design and behaviour of the resulting organisms, and an understanding of why they had evolved in such a way (in order to know which aspects of their design and behaviour were the most important for us to model).⁶ Unfortunately we do not possess such details of the organisms which existed at this stage of evolution on Earth.

I am certainly not the first person to criticise artificial life models on these grounds. For example, Howard Pattee warns that “simulations that are dependent on ad hoc and special-purpose rules and constraints for their mimicry cannot be used to support theories of life” (Pattee, 1988, p.68).

To be fair, Ray does offer a definition of life in his work with *Tierra*. He says “I would consider a system to be living if it is self-replicating, and capable of open-ended evolution” (Ray, 1991, p.372). However, determining necessary and sufficient conditions for a system to be capable of open-ended evolution is half of the problem, and Ray's definition tells us nothing about how we should go about building such a system. This being the case, the definition does not provide an adequate theoretical grounding for *Tierra* and similar models.

A weakness of Ray's definition of life for our present purposes is that it does not define what sorts of environments might support life, or the sorts of ecological interactions which should be available. Now, it is often argued that ecological processes may play a primary role in promoting evolutionary activity and the evolutionary increase of complexity of some organisms; for an overview, see (Taylor, 1999, Section 2.3.1). Furthermore, some of the most spectacular examples of artificial evolution rely upon coevolutionary interactions between organisms, as mentioned in Section 2. This suggests that we should think more carefully about such issues, rather than treating them in the rather ad hoc way that has often been used in the past. This point has been made by Pattee, who says:

“... life must have arisen and evolved in a nonliving milieu. In real life we call this the real physical world. If artificial life exists in a computer, the computer milieu must define an artificial physics ... What is an artificial physics or physics-as-it-could-be? Without principled restrictions this question will not inform philosophy or physics, and will only lead to disputes over nothing more than matters of taste in computational architectures and science fiction.” (Pattee, 1995a, p.29).

The ad hoc feel of *Tierra*-like systems is a direct consequence of this lack of theoretical grounding. The unmanageable parameter space of many of them can also be attributed to this lack of direction. As a result of these weaknesses, even if interesting behaviours are observed

⁶Ray himself recognises these difficulties, but is more optimistic that they can be overcome (Ray, 1991, p.399).

in these systems, we are unlikely to be able to adequately explain them in any general sense without further substantial theorizing and experimentation. It may be that a model of self-replication and open-ended evolution is necessarily somewhat complex, but, even if this is so, the theoretical framework upon which it is built should prescribe the implementational details as much as is practically possible.

3.3 Implicit versus Explicit Encoding

Tierra, as well as many of the other artificial evolutionary systems mentioned in Section 2, can be analysed in terms of von Neumann's work. In this section I analyse Tierra in terms of the various components (e.g. A , ϕ) of his architecture. Specifically, I consider the extent to which these components are explicitly encoded on the evolving individuals themselves, rather than being implicitly encoded in the 'laws of physics' of the environment in which they exist (i.e. the operating system of the platform). Now, as we are interested in the evolution of the self-reproducing individuals in these systems, and as the inheritable information of each individual (i.e. the part which gets passed on from parent to offspring) is contained on the tape ϕ in von Neumann's architecture, I will assume that the tape must be explicitly represented in some fashion, otherwise there would be nothing which could evolve. We can now ask *which parts* of the [$A + B + C + D$] architecture are explicitly encoded on the tape ϕ . Of course, even the behaviour of those parts which are represented on the tape will still to some extent be encoded in the 'laws of physics' of the environment, but I think the analysis is nevertheless worthwhile.

In the case of von Neumann's envisaged implementation of self-reproducing cellular automata, it is clear that all four subcomponents (i.e. A , B , C and D) are very explicitly encoded on the tape $\phi(A + B + C + D)$; the environment in which the automaton exists implicitly encodes only very low-level actions in the form of the local transition rules of individual cells.

I would suggest that the reproducing programs in Tierra and similar systems can also sensibly be analysed in terms of von Neumann's architecture. Before I begin, I would like to make a couple of general points, which might help to reorient the reader to my perspective. Firstly, I believe that the notion of a phenotype fundamentally involves *interaction* with the environment (and that this is the essential distinction between the notions of phenotype and genotype—the latter being an informational concept). When I talk about phenotypes in the following, therefore, and specifically when I talk about the automata A , B , C and D , I am interested in the role these phenotypic structures play—their function—rather than the details of implementation or of how that function is achieved. Secondly, note that the terminology commonly used to describe reproducers in Tierra-like systems is somewhat different to that used for von Neumann's work. Because of the simi-

larity between Tierra-like operating systems and those of standard digital computers, the actions of Tierran reproducers are often referred to as computations rather than constructions, even when a reproducer is in the process of building a new copy of itself. However, this process of reproduction is, of course, central to the Tierra approach, and I believe that this procedure of building a copy of a program in a different part of memory is, in all the relevant details, a process of construction in just the same way as construction processes in von Neumann's cellular automata model. In the following, also remember that von Neumann's general constructing automaton A is the machinery which *interprets* the tape to produce a new machine (phenotype), and the general copying automaton B copies the tape uninterpreted.

At first sight it might seem that there is no separate genetic description of the program in a Tierra-like system. The picture is complicated by the fact that the machinery which interprets the program (i.e. automaton A) does not reside in the same part of the computer in which the program itself is stored. The state information for this machinery—a program's 'virtual CPU' (i.e. the instruction pointer, stacks, registers, etc.)—is generally represented in an independent area of memory to the program's instructions. Furthermore, the actual 'interpreting machinery' of the virtual CPU is encoded in the global operating system provided by the platform, and is in this sense implicit in the program's environment. Additionally, the control automaton C , which controls when the instructions in the program get executed, is also implicit in the part of the operating system which governs mechanisms such as how a program's instruction pointer is updated after the execution of each instruction. All that is left to be explicitly encoded by the program, therefore, is the copying automaton B , and potentially any other arbitrary automaton D .

Now, the instructions which make up the program exist in an unreactive state in the system's random-access memory. It is only when the control automaton C transfers instructions to the interpreting automaton A that they become 'active'. Looked at in this way, we can see that it is the *behaviour* of the program (including looping, jumping, etc.) that is the result of automaton A interpreting the unreactive genetic description. This behaviour is therefore the equivalent to the constructed machine (and the actions it performs—i.e., the phenotype) in von Neumann's design, and the string of instructions residing in the random-access memory (which is normally referred to as the program) is the tape or genetic description of this phenotype. It is perhaps easier to see the distinction if one considers a parallel program, with multiple processes (with different state information) using the same program listing.

I therefore suggest that a self-reproducing program in a Tierra-like system is consistent with von Neumann's architecture. However, as automata A and C are largely implicit in the environment in which the programs reside

(the only explicit representation being the state information in a program's virtual CPU), and are certainly not encoded by the individual programs, we can see that a 'program', in the sense of a string of instructions in the system's random-access memory, corresponds to the tape $\phi(\mathbf{B} + \mathbf{D})$ in von Neumann's scheme.

It is interesting to speculate on what information we might desire to be explicitly encoded on a structure which would be suitable for acting as a robust initial seed for an open-ended, and possibly creative, evolutionary process. I will refer to such a structure as 'proto-DNA'. Now, we would like our proto-DNA to be an indefinite hereditary replicator if it is to be such a seed (Maynard Smith and Szathmáry, 1995). In other words, it should be able to exist in an unlimited number of configurations which retain the ability to reproduce. If the copying process is encoded on the tape itself, then mutations have the potential to disrupt its ability to be reproduced. It would therefore seem desirable that the copying automaton \mathbf{B} of our proto-DNA be largely implicitly encoded in the environment. Note that this would not necessarily prevent a more complicated, and possibly more reliable, explicit copying process \mathbf{B}' later evolving from (but still based upon) the simpler implicit process, as indeed seems to have happened during biological evolution.

If the copying procedure for our proto-DNA is implicitly encoded in the environment, however, any configuration of proto-DNA would, all else being equal, be able to reproduce as well as any other. In other words, there would be no basis for preferentially selecting some configurations over others, and therefore no basis for an evolutionary process. Specific configurations of proto-DNA must therefore have some specific properties that are selectively significant. Models of the origin of life commonly presume that these simple phenotypic properties were things such as increased stability of the molecule, simple control of the local environment, catalytic activity, etc. (e.g. Eigen and Schuster, 1977; Cairns-Smith, 1985; Szathmáry and Demeter, 1987).

At the initial stages of an evolutionary process, however, we would not expect there to be mechanisms for explicitly decoding the proto-DNA; in other words, the interpretation machinery \mathbf{A} is implicit.⁷ This means that particular configurations of proto-DNA should have some specific phenotypic properties (such as the ability to act as catalysts) which can be determined directly from their structure rather than having to be explicitly decoded from the genotype. We could therefore regard the proto-DNA as merely $\phi(\mathbf{D})$, meaning that particular configurations have particular phenotypes associated with them, which are (a) not related to the process of self-reproduction *per se*, and (b) do not require to be decoded by an explicit interpretation automaton \mathbf{A} . Regarding the kinds of

⁷We could, of course, 'hard-wire' explicit interpretation machinery into the system (as in the programming language provided in Tierra), but to do so would inevitably impose restrictions on the evolutionary possibilities available.

simple phenotypes that we might wish to be available to our proto-DNA, some possibilities are suggested by the origin-of-life models mentioned previously, but in general the options seem endless. Graham Cairns-Smith observes:

"It is almost too easy to imagine possible uses for phenotype structures—because the specification for an effective phenotype is so sloppy. A phenotype has to make life easier or less dangerous for the genes that (in part) brought it into existence. There are no rules laid down as to how this should be done." (Cairns-Smith, 1985, p.106).

Proto-DNA with inherent phenotypic properties can therefore serve as a suitable starting point for an open-ended evolutionary process. To digress a little, with regard to the issue of how symbolic information arises in evolution (discussed, for example, by Pattee, 1995b), this requirement ensures that the matter-symbol relationship is inherent in the system from the beginning. The material is selected for its phenotypic properties, but it is its genetic information which is passed on to its offspring. In this situation, it is necessary to assume that by inheriting this genotype, the offspring will also share the phenotypic properties. For example, in a simple RNA-world scenario,⁸ we could imagine that molecules which inherit a particular sequence of bases would adopt a particular three-dimensional structure, which might, say, confer specific catalytic properties (as demonstrated by Zaug and Cech, 1986). We could therefore regard the genetic information (the sequence of bases on the RNA molecule) as a symbolic representation of its phenotypic properties (its catalytic action in this example).

However, there would presumably be a limit on the number of different inherent phenotypic properties these proto-DNA structures might possess. Furthermore, if the proto-DNA is to reliably reproduce it should be a fairly stable molecule, and this requirement further restricts the range of effective phenotypic properties that it might have. If more complicated phenotypes are to arise later on in the evolutionary process, therefore, it appears necessary that a stronger distinction is introduced between genotypes and phenotypes. The biologist C.H. Waddington remarked that:

"... in practice—and perhaps because of a profound law of action-reaction—it is difficult (impossible?) to find a [molecule] which is stable enough to be an efficient store and at the same time reactive enough to be an efficient operator" (Waddington, 1969, p.115).

The advantages of a genotype-phenotype distinction over other forms of reproduction have been discussed by many people; for a review, see (Taylor, 1999, Section 7.2.3).

⁸For references to work on RNA worlds, see Nuño et al. (1995) and Lazcano (1995).

For such a distinction to arise with proto-DNA, we require that it at least has the potential for explicit interpretation machinery A' and control machinery C' to become associated with it. This would involve some form of specific reaction to subsections of information in the proto-DNA, but more work is needed to fully identify how this potential for explicit interpretation might be assured.

3.4 Ability to perform other tasks

In the previous section it was suggested that proto-DNA in its primitive form should not involve much interpretation or control machinery. However, it is important that some specific phenotypic properties are implicitly associated with specific structures (i.e. these properties are apparent without the need for explicit interpretation machinery). Without the ability of individual replicators to have other properties as well as self-reproduction, the evolving system will not be very interesting. Indeed Muller, who, in the early part of this century was the first person to explicitly propose an exclusively evolutionary definition of life, emphasised the importance of this material “affecting other materials and, therewith, its own success in genetic survival” (Muller, 1966, p.512). This picture of individual reproducers affecting other materials reminds us that biological evolution has involved the coevolution of interacting organisations rather than of single, isolated reproducers. As mentioned in Section 3.2, many existing artificial evolutionary systems have concentrated almost exclusively on modelling individuals, with little regard for the principled modelling of interactions between individuals.

Nils Barricelli was well aware of the need for reproducers to perform other tasks when he designed his artificial life platform in the early 1950s. He says “It may appear that the properties one would have to assign to a population of self-reproducing elements in order to obtain Darwinian evolution are of a spectacular simplicity. The elements would only have to: (1) Be self-reproducing and (2) Undergo hereditary changes (mutations) in order to permit evolution by a process based on the survival of the fittest” (Barricelli, 1962, pp.70–71). He goes on to describe a simple discrete one-dimensional model where each cell is either empty or contains an integer number. The numbers reproduce according to the implicit rules of the system, and mutations arise under certain circumstances. This simple model therefore fulfils the fundamental requirements for an evolutionary process. However, as Barricelli notes, this model of evolution “clearly shows that something more is needed to understand the formation of organs and properties with a complexity comparable to those of living organisms. No matter how many mutations occur, the numbers . . . will never become anything more complex than plain numbers” (*ibid.* p.73). Barricelli therefore concentrated on looking for the ‘missing ingredient’.⁹ It should be noted that von Neumann,

⁹His solution was to require that elements could only reproduce in

also, was not so much interested in machines which could only self-reproduce, but rather in machines which could perform other tasks as well (von Neumann (1966) p.92; see also McMullin (1992) pp.174–175).

The preceding arguments are leading us in the direction of requiring a form of proto-DNA which reproduces due to the implicit laws of the environment in which it exists, but which also explicitly specifies some properties which can be selected for or against in an evolutionary process. At this point we might note that artificial evolutionary systems which have just these properties already exist, and indeed their use is widespread; these are the optimisation tools mentioned in Section 2, such as genetic algorithms (e.g. Holland (1975), Goldberg (1989)), genetic programming (e.g. Koza (1992)) and similar techniques. The difference is that we require a system with the potential for a large degree of *intrinsic* adaptation for open-ended evolution, rather than a system where the selection of individuals is determined by an externally-defined fitness function (see Section 2). Intrinsic adaptation is introduced when the *domain of interaction* of the individuals is within the evolving system itself, and the individuals are competing for limited resources. This is in contrast to systems with an explicitly defined fitness function, where the replicators do not directly interact with other replicators. Ray recognised this point himself when discussing the design of artificial life platforms:

“What all of this discussion points to is the importance of imbedding evolving synthetic organisms into a context in which they may interact with other evolving organisms. A counter example is the standard implementations of genetic algorithms in which the evolving entities interact only with the fitness function, and never ‘see’ the other entities in the population. Many interesting behavioral, ecological and evolutionary phenomena can only emerge from interactions among the evolving entities.” (Ray, 1994, Section 11.1).

Similar arguments for proto-DNA with the properties of implicit reproduction and the potential for explicitly-encoded attributes with selective significance have been put forward by Barry McMullin, who points out the connection with Cairns-Smith (1985)’s general model for the origin of terrestrial life based upon inorganic information carriers (McMullin, 1992, p.267).

3.5 Embeddedness in the Arena of Competition and Richness of Interactions

In the preceding sections I have emphasised the importance of the distinction between intrinsic and extrinsic symbiotic association with other elements. While this may indeed be an important aspect of the ‘missing ingredient’, it is extremely doubtful that it is the *only* important aspect.

lection. I will now discuss some issues involved in this distinction in more detail.

An essential requirement for an evolutionary process is that some form of selection mechanism exists, so that some variations of the reproducing entities are favoured over others. The selection mechanism therefore introduces a form of competition between the individual reproducers; they become engaged in a struggle for existence. The presence of such a mechanism implies that, in some form, the individuals coexist in an arena of limited capacity, and that they are competing with their neighbours (either globally or locally) for the right to be there.

An evolutionary system must therefore have an arena of competition of some description, although there are few restrictions on the particular form it should take. All that is required is that it introduces the concept of a resource that is: (a) a vital commodity to individuals in the population; (b) of limited availability; and (c) that individuals can compete for (at either a global or local level). This resource can usually be interpreted as energy, space, matter, or a combination of these.

An issue that arises when considering different evolutionary systems is the extent to which individuals are embedded in this arena of competition. In von Neumann's cellular automata design, individuals are fully embedded—there is no 'hidden' state information (i.e. information which is not embedded in the cellular space itself). The same can be said of the biosphere, at least according to materialism. At the other extreme, individuals in a genetic algorithm (GA) have minimal embeddedness—the arena of competition merely contains place holders for the chromosomes, and the restriction is generally on the number of individuals, regardless of their size (although most GAs have constant-size chromosomes anyway). These two extremes, together with intermediate situations arising in Tierra and Avida,¹⁰ are depicted in Figure 1. Note that individuals in Avida are not really embedded in the arena of competition at all; the two-dimensional environment only holds pointers to the cells, in much the same way as in a GA.¹¹ In Tierra, a program's instructions are embedded in the arena, although each program still has some additional state information (its 'virtual CPU' state). In Avida the fundamental space limitation applies to the number of *programs* that can fit in the arena of competition, whereas in Tierra it applies to the total number of *instructions* contained in all of the programs in the population.

It should be emphasised that this notion of embeddedness is unrelated to the distinction between implicit and

¹⁰Avida is an artificial life platform developed by Chris Adami and colleagues (see <http://www.krl.caltech.edu/avida/>). It is based upon Tierra, but there are some significant differences, especially in the modelling of the environment. For example, individual programs occupy positions in a two-dimensional arena and are only in direct competition for space with their neighbours.

¹¹That is, the two-dimensional environment in which all of the programs coexist is distinct from the one-dimensional memory in which each individual program is stored. Furthermore, in the default settings of Avida, programs cannot read instructions of neighbouring programs, so no parasitism of this nature can emerge.

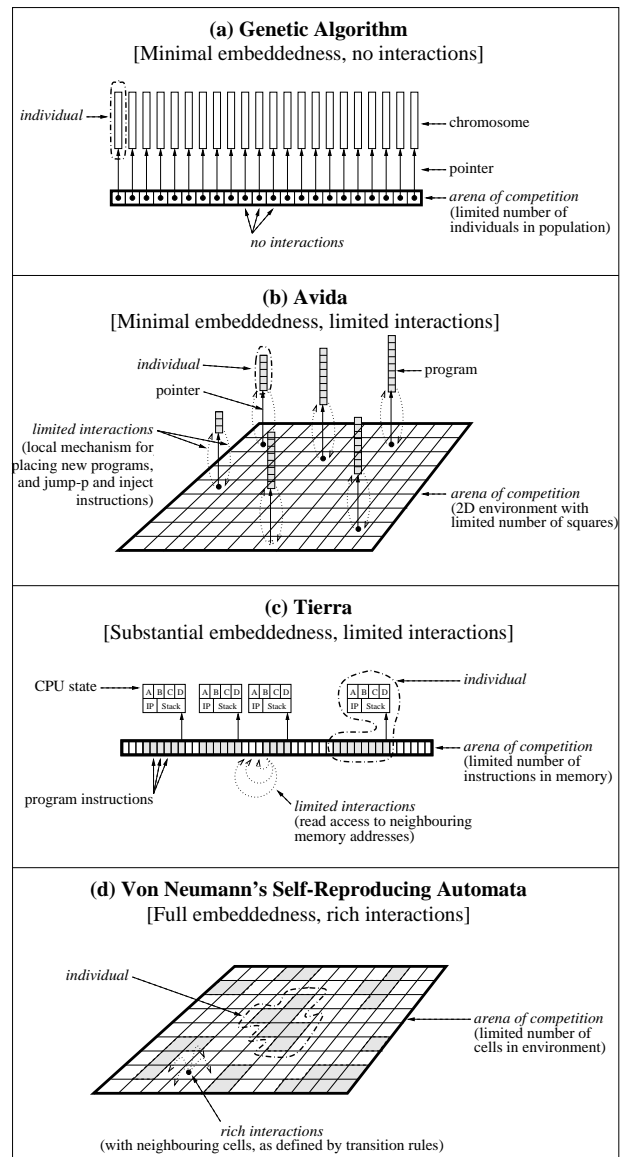


Figure 1: Embeddedness of Individuals and Richness of Interactions in Various Artificial Evolutionary Platforms.

explicit encoding, which concerns the degree to which a process is governed by the environment as opposed to a specific object situated within that environment. The issue of embeddedness concerns the representation of individuals only; it does not (directly) concern the representation of the abiotic environment.

Related to the issue of physical embeddedness is that of how restricted is the range of interactions that are allowed between objects within the arena. In a standard GA, no direct interactions are allowed between chromosomes at all; the continued existence of an individual is decided by the extrinsically-defined selection mechanism. As already mentioned, in the default configuration of Avida programs cannot read the instructions of their neighbours. However, some extra instructions can be enabled to al-

low these sorts of interactions to occur.¹² Although programs in Tierra are embedded in the arena of competition to a much greater extent than they are in Avida, the range of interactions allowed with neighbouring programs is still fairly restricted; programs can read the code of their neighbours, but they cannot directly write to neighbouring memory addresses.

In contrast, von Neumann's cellular automata implementation is far less restrictive; the transition rules of the cellular automata define neighbourhood interactions which occur at the level of individual cells and which therefore do not respect boundaries between individual organisms. This is of course similar to the situation of biological organisms, which have the freedom to interact with their environment in a variety of ways only limited by the laws of physics (although organisms themselves generally evolve mechanisms to restrict such free interaction).

From the point of view of the evolvability of individuals, the more embedded they are, and the less restricted the interactions are, then the more potential there is for the very *structure* of the individual to be modified. Recall that this is one aspect of my definition of *creative* evolution. Sections of the individual which are not embedded in the arena of competition are 'hard-wired' and likely to remain unchanged unless specific mechanisms are included to allow them to change (and the very fact that specific mechanisms are required suggests that they would still only be able to change in certain restricted ways).

Additionally, from an epistemological point of view, Pattee (1995b) points out that symbolic information (such as that contained in an organism's genes) has "no intrinsic meaning outside the context of an entire symbol system as well as the material organization that constructs (writes) and interprets (reads) the symbol for a specific function, such a classification, control, construction, communication ...". He argues that a necessary condition for an organism to be capable of creative open-ended evolution is that it encapsulates this entire self-referent organisation (Pattee refers to this condition as *semantic closure*). From this it follows that organisms should be constructed "with the parts and the laws of an artificial physical world" Pattee (1995a) (p.36).¹³ In other words, the interpretation (phenotype) of the symbolic information (genotype) of an artificial organism should be constructed and act within the artificial physical environment of the system. Additionally, if the system is to model the *origin* of genetic information, then the genotype itself must also be embedded within the environment; that is, the complete semantically-closed organisation—the *entire organism*—must be completely embedded within the physical environment. Pattee's arguments also suggest the need for material, rather than purely formal, models—an issue to be discussed in Section 3.6.

¹²These are the `jump-p` and `inject` instructions.

¹³Although he also stresses that "some epistemic principles must restrict physics-as-it-could-be if it is to be any more than computer games" (Pattee, 1995a).

To end this section, I briefly mention Holland (1995)'s work with the 'Echo' model of complex adaptive systems. Echo possesses many of the features that I have just argued are desirable for a model of open-ended evolution. For example: selection is determined intrinsically by interactions between Echo organisms (or to use Holland's terminology, agents), rather than by an externally-defined fitness function; the process by which agents reproduce is implicitly defined in the Echo operating system rather than being explicitly encoded by individual agents; and the agents are able to perform a variety of phenotypic behaviours; Echo is also designed upon more explicit design considerations than were most earlier artificial life models; the considerations for Echo are based upon a core set of principles which Holland believes are common to all complex adaptive systems. For all these reasons, I believe Echo represents a significant advance. However, the structure of the individual agents—the notion of what it is to be an agent—is still predefined, and the representation of agents is not fully embedded in the arena of competition. Additionally, the interpretation of agent's chromosomes is handled implicitly by the operating system. Now, the system was designed in this way because it is primarily intended as a general model of complex adaptive systems, rather than a specific model of biological evolution. Indeed, the various successful applications of Echo (e.g. Schmitz and Booth, 1996; Hraber and Milne, 1997) testify to the value of the particular way in which the organism and environment structure have been predefined; if no higher-level structure were imposed, it would be difficult to model most complex adaptive systems of interest (e.g. ecologies, economies, etc.).

In the context of open-ended evolution, however, the design still has some shortcomings. The fact that the Echo operating system implicitly interprets the agents' chromosomes means that they can never come to encode anything more than the fixed range of actions (e.g. offence, defence, conditional exchange of resources) predefined by the designer. In *Hidden Order*, Holland discusses how new meaning can arise in a system, but acknowledges that Echo is deficient in this respect (Holland, 1995, p.138). As Pattee has suggested, it is only when an organism's genotype, phenotype, and the interpretation machinery that produces the latter from the former (including the whole developmental process through which an adult phenotype is produced)—that is, the whole semantically closed organisation—is all embedded in the arena of competition that fundamentally new symbolic information can arise in the genome (i.e. the generation of genetic information representing new functional relationships between the organism and its environment). In the discussion of the desirable properties of proto-DNA in Section 3.3, it was suggested that this too would initially be interpreted implicitly. It was, however, stressed that the potential should exist for explicit interpretation machinery to evolve, thereby creating an explicit representation of the whole semantically closed organisation and

allowing the possibility for new symbolic information to arise.

3.6 Materiality

The arguments in the previous sections are bringing us to the fundamental question of how matter is represented in these models. If there is a representational distinction between organisms and the environment in which they exist (which comes about by having a hard-wired organism structure and by restricting ecological interactions), some of the fundamental concepts associated with living beings, such as competition for resources, self-maintenance and so on, become ill-defined. As mentioned in Section 3.2, these kinds of ecological relationships may play a very important role in promoting open-ended (and possibly creative) evolution. It is therefore vital that we consider the issues involved in modelling such relationships if we hope to design artificial systems which have the capacity for open-ended and creative evolution.

One of the tenets of Darwinism is that organisms are engaged in a struggle for existence. However, it is difficult to identify the precise nature of this struggle, as Darwin himself observed. In *The Origin of Species*, he wrote:

“What checks the natural tendency of each species to increase in number is most obscure ... The amount of food for each species of course gives the extreme limit to which each can increase; but very frequently it is not the obtaining food, but the serving as prey to other animals, which determines the average numbers of a species” (Darwin, 1859, pp.119–120).

Thus, an important aspect of the struggle for existence is the obtaining of food not from passive, abiotic sources, but through predator-prey relationships. In the biological realm, the struggle for existence involves organisms *killing* other organisms, because *the very stuff from which they are constructed is a valuable resource of matter and energy*. This competition is therefore very much a matter of life or death.

It may be difficult to identify the precise nature of the struggle for existence, but it seems likely that the numerous forms of competition can be categorised in terms of a small number of fundamental resources (as mentioned in Section 3.5). In the biosphere, a (speculative) list might be: matter, energy, space and information.¹⁴

Tierra-like systems generally do not have any notion of competition for matter. Indeed, they cannot really be said to have a notion of matter at all, in terms of fundamental units from which all structures are built, and which are conserved during reactions. Instead, when a program

¹⁴For example, a virus requires information contained in its host's genome in order to reproduce. This information is more than the matter from which the host's DNA is constructed; it involves a particular ordering of matter.

is writing a copy of itself, it can produce the copied instructions spontaneously rather than first having to collect a copy of the individual instruction from somewhere else in memory. In other words, the individual instructions are represented as states of specific memory locations, rather than as units of matter, as is also the case in von Neumann's cellular automata model; these systems are formal models rather than material models. The only fundamental competition that exists in Tierra is for space (memory) into which to divide. This is allocated at a global level by the Tierran operating system's memory allocation services.

In Tierra, programs are not even really competing for energy (CPU-time), because any number of programs are allowed to execute instructions at each time slice; the limiting factor is how many programs can fit into the available memory. In Avida the situation is somewhat different, as programs can win more CPU-time by successfully performing certain arithmetic challenges presented to them by the environment (Adami and Brown, 1994).

Programs in Tierra *can* act as resources for other programs in another way, by acting as 'library code' which can be read by their neighbours (as happens in the evolution of parasites). In other words, they can act as information resources. However, this is not as strong an ecological interaction as when one organism acts as a resource of matter or energy, in the sense that acting as an information resource is not a direct matter of life or death for the host.

The issue of how energy is represented in these systems is perhaps more controversial. Some would claim that it is essential to model certain fundamental energetic considerations (e.g. Morán et al., 1997; Ruiz-Mirazo et al., 1998). An important point to note is that *all* artificial life platforms have to model energy at the basic level of determining when a component can perform an action (e.g. when a program can execute an instruction, as determined by the system's CPU-time allocation scheme). Without a theoretical grounding, any scheme is just as arbitrary as any other (e.g. the schemes in Tierra and Avida). Ideally, the system's design should be based upon explicit considerations of how energy should be modelled.

Only when one organism can act as a resource of energy and matter for other organisms do ecological concepts such as food webs and trophic levels (which can act as important drives for open-ended evolution) become relevant.

Other advantages of material evolutionary systems over purely formal systems have been suggested by Luis Rocha:

“Material sign systems are not universal and cannot represent anything whatsoever, but this turns out to be their greatest advantage. The price to pay for the universality of formal symbol systems is complete specificity, that is, full description of its components and behaviour. Conversely, material sign systems are built over certain building blocks which

do not need a description. For instance, DNA does not need to encode anything but aminoacid chains, there is no need to include in genetic descriptions information regarding the chemical constituents of aminoacids not instructions on how to fold an aminoacid chain—folding comes naturally from the dynamical self-organization of aminoacid chains.” (Rocha, 1998).

In other words, the genome does not have to encode information about every aspect of the organism’s phenotype, because some features will just fall into place ‘for free’, due to the self-organizational properties of the constituent matter. This may significantly ease the problem of evolving complex phenotypes.

4 A Full Specification for an Open-Ended Evolutionary Process

Perhaps the most important point to arise from the preceding discussion is that processes such as self-reproduction operate *within an environment* rather than in isolation. The properties of this environment, and the ways in which evolving entities may interact with it (and with each other), fundamentally influence the evolutionary process.

Reflecting upon the significance of his work on evolution, and in particular on his demonstration of the possibility of machines which could build modified copies of themselves, von Neumann said “It is clear that this is a step in the right direction, but it is also clear that it requires considerable additional analyses and elaborations to become really relevant” (von Neumann, 1966, p.131).

It has long been recognised that chief among these additional analyses and elaborations is the incorporation of the evolutionary process into a broader framework that also considers the properties of the environment. Holland has emphasised that the study of adaptation “involves the study of both the adaptive systems and its environment. In general terms, it is a study of how systems can generate procedures enabling them to adjust efficiently to their environments” (Holland, 1962, p.299). Moreover, Conrad (1988) stresses that “the characterization of the substrate is of such immense importance for the effectiveness of evolution” (p.304).

Studies of evolution *in vitro*, such as Orgel (1979)’s experiments with evolving RNA sequences using the viral enzyme $Q\beta$ replicase, have also demonstrated the need for a better theoretical understanding of these issues. Maynard Smith explains:

“More or less independently of the starting point ... the end point is a rather small molecule, some 200 bases long, with a particular sequence and structure that enable it to be replicated particularly rapidly. In this

simple and well-defined system, natural selection does not lead to continuing change, still less to anything that could be recognized as an increase in complexity: it leads to a stable and rather simple end point. This raises the following simple, and I think unanswered, question: What features must be present in a system if it is to lead to indefinitely continuing evolutionary change?” (Maynard Smith, 1988, p.221).

The question raised by Maynard Smith is exactly the one of interest in this paper: What sort of system (in terms of individuals, interactions and environments) will give rise to an open-ended, and possibly creative, evolutionary process?

4.1 Waddington’s Paradigm for an Evolutionary Process

A characterisation of a process which might be capable of supporting open-ended evolution was proposed by C.H. Waddington 30 years ago (Waddington, 1969). He went as far as to call this characterisation a new paradigm under which biological evolution should be studied. This paradigm is of particular interest because it provides a general characterisation of the individuals involved, of how they interact, and of the kind of environment in which they reside. To my knowledge, little work has been devoted to exploring Waddington’s proposal, probably because of the difficulties in capturing it fully with an analytical model (the traditional approach of theoretical biology). However, it is formulated in a way which makes it particularly amenable to synthetic (artificial life) modelling, and is therefore an ideal starting place for developing a better theoretical understanding of open-ended evolution within an artificial life framework.

Waddington describes a replicator as “a material structure \mathcal{P} with a characteristic \mathcal{Q} such that the presence of \mathcal{P} with \mathcal{Q} produces \mathcal{Q} in a range of materials \mathcal{P}_i under circumstances \mathcal{E}_j ” (*ibid.* p.115). In other words, \mathcal{Q} is the characteristic of a structure \mathcal{P} which is inherited when \mathcal{P} is replicated— \mathcal{Q} is the genetic component of \mathcal{P} . The overall scenario is summarised as follows:

“The complete paradigm must therefore include the following items: A genetic system whose items (\mathcal{Q} s) are not mere information, but are algorithms or programs which produce phenotypes (\mathcal{Q}^* s). There must be a mechanism for producing an indefinite variety of new \mathcal{Q}^* s, some of which must act in a radical way which can be described as ‘rewriting the program’. There must also be an indefinite number of environments, and this is assured by the fact that the evolving phenotypes are components of environments for their own or other species. Further, some at

least of the species in the evolving biosystem must have means of dispersal, passive or active, which will bring them into contact with the new environments (under these circumstances, other species may have the new environments brought to them). These environments will not only exert selective pressure on the phenotypes, but will also act as items in programs, modifying the epigenetic processes with which the Q s become worked out into [Q^* s].” (Waddington, 1969, p.120).¹⁵

This general characterisation raises some important issues. For example, the requirement that Q s act not only as information but also as algorithms—that they must act as operators as well as operands—locates the relationship between genotype and phenotype at the very heart of the paradigm. (The same requirement was suggested for proto-DNA, in Section 3.4.) Waddington points out that the open-ended nature of his model relies on the fulfillment of two conditions: (1) that \mathcal{E}_j is an infinite-numbered set; and (2) that there are sufficient Q s to provide Q^* s suitable for an infinite sub-set of \mathcal{E}_j s.

The first condition is satisfied by the fact that Q^* s are components of \mathcal{E}_j s. A vital direction for future research is the investigation of the different sorts of ways in which Q^* s could be components of \mathcal{E}_j s (i.e. how organisms form part of the environment experienced by other organisms), and the evolutionary consequences of such choices.

Of the other condition, Waddington says that “the second requirement, that the available genotypes must be capable of producing phenotypes which can exploit the new environments, requires some special provision of a means of creating genetic variation . . . It is important to emphasize that the new genetic variation must not only be novel, but must include variations which make possible the exploration of environments which the population previously did not utilize . . . It is not sufficient to produce new mutations which merely insert new parameters into existing programmes; they must actually be able to rewrite the programme” (*ibid.* pp.116–118). The distinction Waddington is making here is closely related to my distinction between creative evolution and (merely) open-ended evolution. Another important direction for future research is to explore how this second condition can be satisfied.

It is worth mentioning that some existing artificial evolutionary systems, such as Barricelli (1963)’s studies with evolving game strategies, Conrad and Pattee (1970)’s model, and Holland’s α -Universes (Holland, 1976), do have the notion of emergent operators (phenotypes). However, these phenotypes generally have a limited range of action, thereby preventing the systems from engaging in truly open-ended evolutionary processes.

Now, the requirement in systems capable of open-ended evolution that individual reproducers have selectively significant phenotypic properties, on top of the ability to reproduce, has already been discussed (see Section 3.4). However, it may turn out that the fulfillment of Waddington’s second condition would require reproducing structures to possess not just one, but *multiple* phenotypic properties, possibly of different functional modalities (e.g. catalysis, light sensitivity, motility, etc.). Maynard Smith has observed that “it seems to be a general feature of evolution that new functions are performed by organs which arise, not *de novo*, but as modifications of pre-existing organs” (Maynard Smith, 1986, p.46). This principle could potentially solve the problem raised by Waddington and Pattee, of how new measuring devices arise during evolution (another aspect of my definition of *creative evolution*): a structure with multiple properties might originally be selected for one of these properties, but it might later turn out (quite accidentally) that some of its other properties also confer (unrelated) adaptive advantages upon the bearer of that structure. In such a scenario, an organism which duplicated this structure might have an adaptive advantage over those possessing a single copy, because each structure could be optimised for a single property. In this way, the organism can acquire fundamentally new phenotypic properties. Regarding the issues discussed in Section 3.6, note that these considerations would seem to require a notion of materiality rather than a purely formal model.

This perspective may bring some light to bear upon creative evolution, but it also opens up a whole range of new problems relating to the modelling of multiple, and mostly (initially at least) irrelevant, properties of objects. Such questions require much more investigation, but existing work reported in the biological literature on multifunctional enzymes may be helpful (e.g. Kacser and Beeby, 1984).

5 Summary

In this chapter I have discussed the concept of open-ended evolution, and the introduction of fundamental novelty during evolution (i.e. creative evolution). Creativity is more subtle than open-ended evolution, and involves issues such as the emergence of symbolic information, and the evolution of new measuring instruments. I have analysed some existing artificial evolutionary platforms in terms of their ability to support open-ended and creative evolutionary processes. The discussion emphasises that existing models have generally concentrated on the representation of individuals, and that explicit theoretical considerations concerning the design of the environment (including the issue of how individuals form part of the environment experienced by others, the degree of implicit versus explicit encoding of processes, and the issue of material versus formal models), and of the sorts of in-

¹⁵In the original paper, the final word of this paragraph appears as Q 's rather than Q^* s. This is fairly clearly a typographical error.

teractions allowed between individuals and their environment, have often been lacking. I have also discussed the desirable properties of proto-DNA—a hypothetical structure which might be suitable to act as a seed for an open-ended, and creative, evolutionary process. I suggested that the capacity of this proto-DNA to reproduce should not be easily disrupted by mutations, and therefore that the reproduction process should be implicitly encoded in the environment rather than explicitly encoded on individuals. This led to a discussion of the sorts of phenotypic properties that should be associated with specific proto-DNA structures, on top of their ability to reproduce. In addition, the environment in which the proto-DNA exists should allow unrestricted interactions between individuals, and the representation of individuals should be fully embedded within the arena of competition of the system, so as not to limit the structure's evolutionary potential. I have suggested that the development of material models, as opposed to purely formal ones, may be a useful avenue to explore; in particular, the modelling of matter with phenotypic properties in a number of different modalities. Throughout the paper I have highlighted various open questions relating to these issues which need to be addressed by future research. In Section 4 I described a paradigm suggested by Waddington, which might represent a suitable starting place for a more unified and productive exploration of these issues using synthetic (artificial life) modelling techniques.

Acknowledgements

I would like to thank John Hallam, Nick Barton, John Holland, Howard Pattee and Moshe Sipper for useful discussions on the issues addressed in this chapter.

References

- Chris Adami and C. Titus Brown. Evolutionary learning in the 2D artificial life system 'Avida'. In R Brooks and P Maes, editors, *Artificial Life IV*, pages 377–381. The MIT Press, 1994.
- W. Brian Arthur. On the evolution of complexity. In G. Cowan, D. Pines, and D. Meltzer, editors, *Complexity: Metaphors, Models and Reality*, volume XIX of *SFI Studies in the Sciences of Complexity*, pages 65–81, Reading, MA, 1994. Addison-Wesley.
- W. Aspray and A. Burks, editors. *Papers of John von Neumann on Computing and Computer Theory*. MIT Press, 1987.
- Nils Aall Barricelli. Symbiogenetic evolution processes realized by artificial methods. *Methodos*, 9(35-36), 1957.
- Nils Aall Barricelli. Numerical testing of evolution theories. Part I. Theroetical introduction and basic tests. *Acta Biotheoretica*, XVI(1/2):69–98, 1962.
- Nils Aall Barricelli. Numerical testing of evolution theories. Part II. Preliminary tests of performance. Symbiogenesis and terrestrial life. *Acta Biotheoretica*, XVI(3/4):99–126, 1963.
- Mark A. Bedau. Four puzzles about life. *Artificial Life*, 4(2):125–140, 1998.
- A.G. Cairns-Smith. *Seven Clues to the Origin of Life*. Cambridge University Press, 1985.
- Michael Conrad. The price of programmability. In R. Herken, editor, *The Universal Turing Machine: A Half-Century Survey*, pages 285–307. Oxford University Press, 1988.
- Michael Conrad and H.H. Pattee. Evolution experiments with an artificial ecosystem. *Journal of Theoretical Biology*, 28:393–409, 1970.
- Charles Darwin. *The Origin of Species*. John Murray, London, 1859. (Reprinted in Penguin Classics, 1985. Any page numbers given in the text refer to this reprint.)
- Manfred Eigen and Peter Schuster. The hypercycle: A principle of natural self-organization. *Die Naturwissenschaften*, 64(11):541–565, 1977.
- Dario Floreano, Stefano Nolfi, and Francesco Mondada. Competitive co-evolutionary robotics: From theory to practice. In R. Pfeifer, B. Blumberg, J.-A. Meyer, and S.W. Wilson, editors, *From Animals to Animats 5: Proceedings of the Fifth International Conference of the Society for Adaptive Behavior*, Cambridge, MA., 1998. MIT Press.
- David E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading, MA, 1989.
- W. Daniel Hillis. Co-evolving parasites improve simulated evolution as an optimization procedure. *Physica D*, 42:228–234, 1990.
- John H. Holland. Outline for a logical theory of adaptive systems. *Journal of the Association for Computing Machinery*, 9(3):297–314, 1962. (Reprinted in *Essays on Cellular Automata*, A.W. Burks (ed.), University of Illinois Press, 1970. Any page numbers given in text refer to this reprint.)
- John H. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, 1975.

- John H. Holland. Studies of the spontaneous emergence of self-replicating systems using cellular automata and formal grammars. In A. Lindenmayer and G. Rozenberg, editors, *Automata, Languages, Development*, pages 385–404. North-Holland, New York, 1976.
- John H. Holland. *Hidden Order: How Adaptation Builds Complexity*. Addison-Wesley/Helix Books, 1995.
- Peter T. Hraber and Bruce T. Milne. Community assembly in a model ecosystem. *Ecological Modelling*, 103:267–285, 1997.
- Henrik Kacser and Richard Beeby. Evolution of catalytic proteins. *Journal of Molecular Evolution*, 20:38–51, 1984.
- John R. Koza. *Genetic Programming: on the Programming of Computers by Means of Natural Selection*. MIT Press, 1992.
- Antonio Lazcano. Prebiotic chemistry, artificial life and complexity theory: What do they tell us about the origin of biological systems? In F. Morán, A. Moreno, J.J. Merelo, and P. Chacón, editors, *Advances in Artificial Life: Third European Conference on Artificial Life*, Lecture Notes in Artificial Intelligence, pages 105–115, Berlin, 1995. Springer.
- John Maynard Smith. *The Problems of Biology*. Oxford University Press, 1986.
- John Maynard Smith. Evolutionary progress and levels of selection. In M.H. Nitecki, editor, *Evolutionary Progress*, pages 219–230. University of Chicago Press, 1988.
- John Maynard Smith and Eörs Szathmáry. *The Major Transitions in Evolution*. W.H. Freeman, Oxford, 1995.
- Barry McMullin. *Artificial Knowledge: An Evolutionary Approach*. PhD thesis, Department of Computer Science, University College Dublin, 1992. URL: <ftp://ftp.eeng.dcu.ie/pub/alife/bmcm.phd/>.
- Geoffrey F. Miller and Dave Cliff. Protean behavior in dynamic games: Arguments for the co-evolution of pursuit-evasion tactics. In D. Cliff, P. Husbands, J.-A. Meyer, and S. Wilson, editors, *From Animals to Animats 3, Proceedings of the 3rd International Conference on Simulation of Adaptive Behavior (SAB'94)*, pages 411–420. MIT Press/Bradford Books, 1994.
- Edward F. Moore. Machine models of self-reproduction. In *Proceedings of Symposia in Applied Mathematics*, volume 14, pages 17–33. The American Mathematical Society, 1962. (Reprinted in *Essays on Cellular Automata*, A.W. Burks (ed.), University of Illinois Press, 1970).
- Federico Morán, Alvaro Moreno, Eric Minch, and Francisco Montero. Further steps towards the realistic description of the essence of life. In C.G. Langton and K. Shimohara, editors, *Artificial Life V: Proceedings of the Fifth International Workshop on the Synthesis and Simulation of Living Systems*, pages 216–224, Cambridge, MA, 1997. MIT Press.
- H.J. Muller. The gene material as the initiator and the organizing basis of life. *The American Naturalist*, 100(915):493–517, 1966.
- John Myhill. The converse of Moore's Garden-of-Eden theorem. In *Proceedings of the American Mathematical Society*, volume 14, pages 658–686, 1963. (Reprinted in *Essays on Cellular Automata*, A.W. Burks (ed.), University of Illinois Press, 1970).
- Juan C. Nuño, Pablo Chacón, Alvaro Moreno, and Federico Morán. Compartmentation in replicator models. In F. Morán, A. Moreno, J.J. Merelo, and P. Chacón, editors, *Advances in Artificial Life: Third European Conference on Artificial Life*, Lecture Notes in Artificial Intelligence, pages 116–127, Berlin, 1995. Springer.
- L.E. Orgel. Selection in vitro. *Proceedings of the Royal Society, London, B*, 205:435–442, 1979.
- Norman H. Packard. Intrinsic adaptation in a simple model for evolution. In C.G. Langton, editor, *Artificial Life*, volume VI of *Santa Fe Institute Studies in the Sciences of Complexity*, pages 141–155, Reading, MA, 1988. Addison-Wesley.
- H.H. Pattee. Simulations, realizations, and theories of life. In C. Langton, editor, *Artificial Life*, volume VI of *Santa Fe Institute Studies in the Sciences of Complexity*, pages 63–77, Reading, MA, 1988. Addison-Wesley. Reprinted in "The Philosophy of Artificial Life", M.A. Boden (ed.), Oxford University Press, 1996.
- H.H. Pattee. Artificial life needs a real epistemology. In F. Morán, A. Moreno, J.J. Merelo, and P. Chacón, editors, *Advances in Artificial Life: Third European Conference on Artificial Life*, Lecture Notes in Artificial Intelligence, pages 23–38, Berlin, 1995a. Springer.
- H.H. Pattee. Evolving self-reference: Matter, symbols, and semantic closure. *Communication and Cognition—Artificial Intelligence*, 12(1–2):9–28, 1995b.
- Umberto Pesavento. An implementation of von Neumann's self-reproducing machine. *Artificial Life*, 2(4):337–354, 1995.
- Thomas S. Ray. An approach to the synthesis of life. In C.G. Langton, C. Taylor, J.D. Farmer, and S. Rasmussen, editors, *Artificial Life II*, pages 371–408. Addison-Wesley, Redwood City, CA, 1991.

- Thomas S. Ray. An evolutionary approach to synthetic biology: Zen and the art of creating life. *Artificial Life*, 1(1/2):195–226, 1994.
- Thomas S. Ray. Evolving complexity. *Artificial Life and Robotics*, 1:21–26, 1997.
- Luis Mateus Rocha. Selected self-organization and the semiotics of evolutionary systems. In S. Salthe, G. Van de Vijver, and M. Delpo, editors, *Evolutionary Systems: Biological and Epistemological Perspectives on Selection and Self-Organization*, pages 341–358. Kluwer Academic Publishers, 1998.
- Kepa Ruiz-Mirazo, Alvaro Moreno, and Federico Morán. Merging the energetic and the relational-constructive logic of life. In C. Adami, R. Belew, H. Kitano, and C. Taylor, editors, *Proceedings of Artificial Life VI*, pages 448–451, MA, 1998. MIT Press.
- Oswald J. Schmitz and Ginger Booth. Modeling food web complexity: the consequence of individual-based spatially explicit behavioral ecology on trophic interactions. *Evolutionary Ecology*, 11:379–398, 1996.
- Karl Sims. Evolving 3D morphology and behavior by competition. In R. Brooks and P. Maes, editors, *Proceedings of Artificial Life IV*, pages 28–39, Cambridge, MA, 1994. MIT Press.
- Eörs Szathmáry and László Demeter. Group selection of early replicators and the origin of life. *Journal of Theoretical Biology*, 128:463–486, 1987.
- Tim Taylor. The COSMOS artificial life system. Working Paper 263, Department of Artificial Intelligence, University of Edinburgh, 1997. Available from <http://www.dai.ed.ac.uk/~daidb/people/homes/timt/papers/>.
- Tim Taylor. *From Artificial Evolution to Artificial Life*. PhD thesis, Division of Informatics, University of Edinburgh, 1999.
- Alan M. Turing. On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London Mathematical Society, Series 2*, 42:230–265, 1936.
- L. Van Valen. A new evolutionary law. *Evolutionary Theory*, 1:1–30, 1973.
- John von Neumann. *The Theory of Self-Reproducing Automata*. University of Illinois Press, Urbana, Ill., 1966.
- C.H. Waddington. Paradigm for an evolutionary process. In C.H. Waddington, editor, *Towards a Theoretical Biology*, volume 2, pages 106–128. Edinburgh University Press, 1969.
- A.J. Zaugg and T.R. Cech. The intervening sequence RNA of *Tetrahymena* is an enzyme. *Science*, 231:470–475, 1986.