Creativity in Evolution: Individuals, Interactions and Environments*

Tim Taylor
Institute of Perception, Action and Behaviour
Division of Informatics, University of Edinburgh
timt@dai.ed.ac.uk

Abstract

This paper 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 paper 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 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 (for example, Holland (1975), Koza (1992)). 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 (for example, Hillis (1990), Sims (1994), Miller and Cliff (1994), Floreano et al. (1998)), where the success of organisms in one population depends upon the success of organisms in another, coevolving population. However, these 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 all together with the notion of modelling evolution towards any sort of high-level goal (e.g. Barricelli (1957), Conrad and Pattee (1970), Packard (1988), Ray (1991),

[0] This paper is an abbreviated version of certain sections of Taylor (1999).

Adami and Brown (1994), Holland (1995)). In these systems, individuals are competing for one or more shared resource (e.g. memory or CPU-time), and the fact that these resources are limited induces natural (intrinsic) selection for those individuals that outcompete 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 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), 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 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. 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. These considerations raise a number of issues, including:

- How does symbolic information arise during an evolutionary process? In other words, how do individuals come to form representations of aspects of their environment?

- How do fundamentally new measuring instruments evolve (i.e. phenotypes that can measure previously unmonitored aspects of the environment)? The significance of this question has been discussed in detail by Pattee (1988).

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.

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 analyse self-reproduction in a number of artificial evolutionary systems in terms of von Neumann's proposed architecture. I then 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.[2] 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, $\mathbf{A}$, which could read a description $\phi(\mathbf{X})$ of another machine, $\mathbf{X}$,

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

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

and build a copy of $\mathbf{X}$ from this description:

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

(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, $\mathbf{B}$, which could copy the instruction tape:

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

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

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

Now, if we choose $\mathbf{X}$ to be $(\mathbf{A} + \mathbf{B} + \mathbf{C})$, then the end result is:

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

This complete machine plus tape, $[\mathbf{A} + \mathbf{B} + \mathbf{C} + \phi(\mathbf{A} + \mathbf{B} + \mathbf{C})]$, 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 $\mathbf{D}$ to the input tape. This gives us:

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

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 $\mathbf{D}$ is changed, but that of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ 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:

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

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

3.2 Implicit versus Explicit Encoding

Many of the artificial evolutionary systems mentioned in Section 2 can be analysed in terms of von Neumann's work. In this section I analyse some of them in terms of the various components 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, 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 $[\mathbf{A} + \mathbf{B} + \mathbf{C} + \mathbf{D}]$ 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 self-reproducing cellular automata, it is clear that all four subcomponents (i.e. $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$) are very explicitly encoded on the tape $\phi(\mathbf{A} + \mathbf{B} + \mathbf{C} + \mathbf{D})$; 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.

The reproducing programs in Tierra (Ray, 1991) and similar systems can also be analysed in terms of von Neumann's architecture. At first sight it might seem that there is no separate genetic description of the program in a system such as Tierra. The picture is complicated by the fact that the machinery which interprets the program (i.e. automaton $\mathbf{A}$) 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 $\mathbf{C}$, which controls when the instructions in the program are 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 $\mathbf{B}$, and potentially any other arbitrary automaton $\mathbf{D}$.

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 $\mathbf{C}$ transfers instructions to the interpreting automaton $\mathbf{A}$ that they become 'active'. Looked at in this way, we can see that it is the *behaviour* of the program (including looping, jumping around the code, etc.) that is the result of automaton $\mathbf{A}$ interpreting the unreactive genetic description. This behaviour, or computation, is therefore the equivalent to the constructed machine, or phenotype, in von Neumann's design.[3] The string of instructions residing in the random-access memory (which is normally referred to as the program) can now been seen as nothing more than the tape or genetic description of this phenotype.

A self-reproducing program in a Tierra-like system is therefore consistent with von Neumann's architecture. However, as automata $\mathbf{A}$ and $\mathbf{C}$ 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 the '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. Notice that with this design the 'genetic code' which maps the genotype $\phi(\mathbf{B}+\mathbf{D})$ to the phenotype $[\mathbf{B}+\mathbf{D}]$ cannot itself evolve, because the interpretation automaton $\mathbf{A}$ is not encoded on the tape.

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. 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 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).

If more complicated phenotypes are to arise later on in the evolutionary process, however, we require that the proto-DNA at least has the potential for explicit interpretation machinery $\mathbf{A}'$ and control machinery $\mathbf{C}'$ 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.

3.3 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). Furthermore, it was suggested in the previous section that the proto-DNA should also have the potential

[3] Indeed, for organisms in *any* kind of evolving system, the notion of a phenotype fundamentally involves behaviour, in the form of interaction with the (biotic and abiotic) environment.

to be explicitly interpreted. 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).

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,[4] 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, the question of how explicit interpretation machinery evolves is more complicated (as mentioned in the previous section).

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 ('trivial reproduction' in the common use of the phrase), 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'.[5] It should be noted that von Neumann, 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.

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 original of terrestrial life based upon inorganic information carriers (McMullin, 1992, p.267).

3.4 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 selection. 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

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

[5] His solution was to require that elements could only reproduce in 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.

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). If one believes in materialism, the same can be said of the biosphere. 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 Cosmos[6] and Tierra, are depicted in Figure 1. Note that individuals in Cosmos 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.[7] In Tierra, a program's instructions are embedded in the arena, although each program still has some additional state information.

It should be emphasised that this notion of embeddedness is unrelated to the distinction between implicit and 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. In Cosmos, programs cannot directly interact with their neighbours, but they can exchange messages and energy tokens via the local environment. Although programs in Tierra are embedded in the arena of competition to a much greater extent than they are in Cosmos, 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.

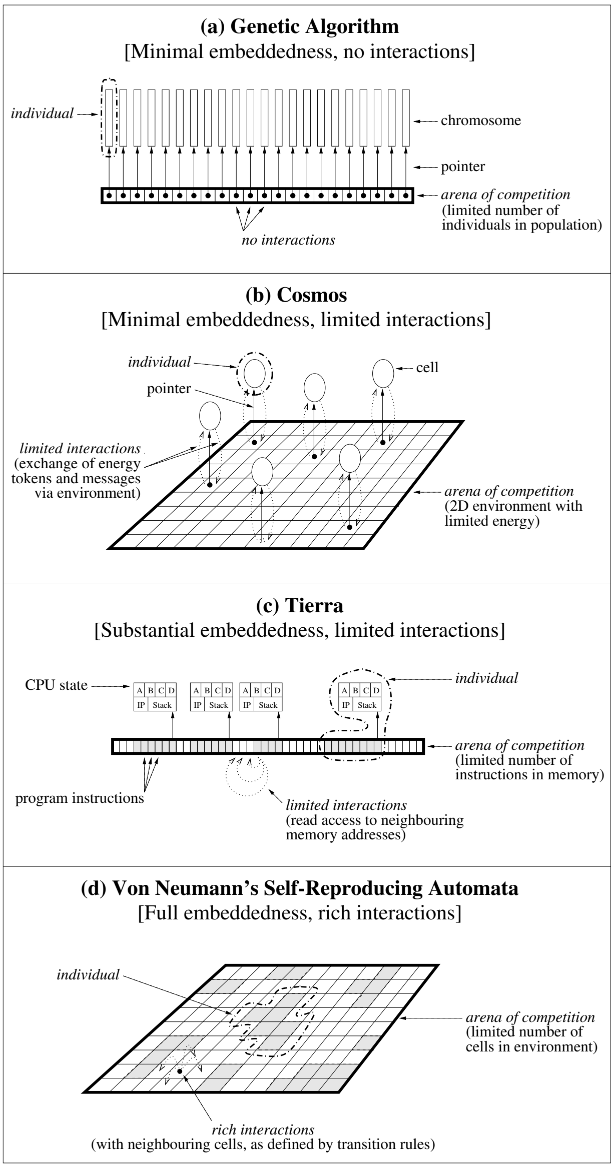
(a) Genetic Algorithm
[Minimal embeddedness, no interactions]
individual
chromosome
pointer
arena of competition
(limited number of individuals in population)
no interactions
(b) Cosmos
[Minimal embeddedness, limited interactions]
individual
cell
pointer
limited interactions
(exchange of energy tokens and messages via environment)
arena of competition
(2D environment with limited energy)
(c) Tierra
[Substantial embeddedness, limited interactions]
CPU state
A B C D
IP Stack
individual
arena of competition
(limited number of instructions in memory)
program instructions
limited interactions
(read access to neighbouring memory addresses)
(d) Von Neumann's Self-Reproducing Automata
[Full embeddedness, rich interactions]
individual
arena of competition
(limited number of cells in environment)
rich interactions
(with neighbouring cells, as defined by transition rules)

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

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. 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

[6] Cosmos is an artificial life platform written by the author (Taylor, 1997). The general design was influenced by Tierra, but there are some fairly significant differences between the two systems. One difference, which it shares with Avida (Adami and Brown, 1994), is that individuals occupy positions in a two-dimensional environment.

[7] The same applies to similar artificial life platforms with two-dimensional environments, such as Avida (Adami and Brown, 1994).

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).[8] 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.

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 in 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. 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 (see Section 3.4), it is only when an organism's genotype, phenotype, and the interpretation machinery that produces the latter from the former, are all embedded in the arena of competition that fundamentally new symbolic information can arise in the genome (thereby permitting truly open-ended evolution). In the discussion of the desirable properties of proto-DNA in Section 3.2, 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 (although how this potential might be assured is an open question).

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 Qβ 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 continu-

[8] Although 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).

> ing 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). 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 're-writing 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 $\mathcal{Q}$s become worked out into [$\mathcal{Q}^*$s]." Waddington (1969) (p.120).[9]

This general characterisation raises some important issues. For example, the requirement that $\mathcal{Q}$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 3.3.) 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 $\mathcal{Q}$s to provide $\mathcal{Q}^*$s suitable for an infinite sub-set of $\mathcal{E}_j$s.

The first condition is satisfied by the fact that $\mathcal{Q}^*$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 $\mathcal{Q}^*$s could be components of $\mathcal{E}_j$s, 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).

Another important direction for future research is to explore how this second condition can be satisfied. Providing the $\mathcal{Q}$s with access to sufficient processes to ensure (something close to) universal construction will undoubtedly be part of the solution. This does not necessarily mean that each $\mathcal{Q}^*$ has to be a universal constructor, but they should at least have access to a basic set of operations to give the set of all $\mathcal{Q}$s the ability to construct a sufficient set of $\mathcal{Q}^*$s. This task may be related to the ability to perform universal computation, which depends on the combination and conditional iteration of a simple set of operations (e.g. Gandy, 1988), although the spatial aspect of construction is an extra complication.

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.

Returning to Waddington's paradigm, notice that his second condition for open-ended evolution is more subtle than that of universal construction alone. A full analysis

[9] In the original paper, the final word of this paragraph appears as $\mathcal{Q}'$s rather then $\mathcal{Q}^*$s. This is fairly clearly a typographical error.

of this condition would also involve the question of how phenotypes which are, in some sense, fundamentally new may be introduced into the population to take advantage of new environments. This question is related to the distinction between creativity and open-ended evolution, discussed in Section 2.

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 3.3). 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 (or fundamentally new phenotypes) arise during 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. This perspective may bring some light to bear upon the evolution of fundamental innovations, 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 paper 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 spatial structure, the issue of how individuals form part of the environment experienced by others, and the degree of implicit versus explicit encoding of processes), and of the sorts of interactions 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. 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 paper.

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. Theoretical 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.

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.

Robin Gandy. The confluence of ideas in 1936. In R. Herken, editor, *The Universal Turing Machine: A Half-Century Survey*, pages 55–111. Oxford University Press, 1988.

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.

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.

H.J. Muller. The gene material as the initiator and the organizing basis of life. *The American Naturalist*, 100(915):493–517, 1966.

Juan C. Nuño, Pablo Chacón, Alvaro Moreno, and Federico Morán. Compartimentation 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.

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. Evolving complexity. *Artificial Life and Robotics*, 1: 21–26, 1997.

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. (Submitted for examination, March 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. Zaug and T.R. Cech. The intervening sequence RNA of *Tetrahymena* is an enzyme. *Science*, 231:470–475, 1986.