in Artificial Life VIII, *Standish, Abbass, Bedau (eds)(MIT Press) 2002. pp 49–52* 1

Von Neumann's Machine in the Shell: Enhancing the Robustness of Self-Replication Processes

Hiroki Sayama

Department of Human Communication, University of Electro-Communications
1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan
sayama@hc.uec.ac.jp

Abstract

One of the grand challenges in artificial life is to create an artificial system that demonstrates self-replication and evolution in the real physical world (Bedau *et al.* 2000). This question was obviously initiated by von Neumann's seminal work on the self-replicating and evolving automaton that was presented more than five decades ago (von Neumann 1951; 1966). Unlike his well known idea of universal computer, however, von Neumann's universal constructor has never been subject to physical implementation (Friedman 1996). This is mainly due to the fragility of its mechanisms against perturbations that are unavoidable in more realistic kinematic settings. In this brief concept paper, I present a simple, but substantial, I believe, idea of enhancing the robustness of the self-replication processes, by introducing an additional subsystem that constructs a "workplace" prior to automaton construction. Workplaces are assumed to be solid structures that can be easily assembled under perturbations and can rigidly hold other components during the automaton construction processes. Similar kind of strategies are found to be prevalent in biological organisms, suggesting the effectiveness of the presented idea for the realization of kinematic models of self-replication and evolution.

Introduction

It is well acknowledged that von Neumann, a great mathematician/physicist who may be best known as a father of the programmable architecture adopted in today's computers, is also one of the pioneers of artificial life for his seminal work on the self-replicating and evolving automaton (von Neumann 1951; 1966). In his latest years, von Neumann challenged an empirical rule that was believed in engineering disciplines that the complexity of products is always smaller than that of the manufacturing machines that produce them. To grapple with this problem, he developed a theoretical discussion and concluded that a counterexample to this rule could exist if a system is made of the following (von Neumann 1951):

A: a universal constructor that creates any arbitrary structure by referring to a static "description tape".

B: a tape duplicator that makes a copy of the description tape.

C: a controller that reads the description tape and passes the written information to the above A and B and appropriately coordinates their behaviors.

I_{A+B+C}: a description tape that specifies how to construct the entire system itself.

These are symbolically written as

$$A + I_X \quad \rightarrow \quad A + I_X + X, \qquad (1)$$
$$B + I_X \quad \rightarrow \quad B + I_X + I_X, \qquad (2)$$

$$
\begin{aligned}
(A + B + C) &+ I_X \\
&\rightarrow (A + B + C) + I_X + X + I_X,
\end{aligned} \qquad (3)
$$

$$
\begin{aligned}
(A + B + C) &+ I_{A+B+C} \\
&\rightarrow (A + B + C) + I_{A+B+C} \\
&\quad + (A + B + C) + I_{A+B+C},
\end{aligned} \qquad (4)
$$

where the last form represents a self-replicating process. More importantly, this model also captures the capability of the evolutionary growth of complexity (McMullin 2000): If the description tape happens to contain an additional content that does not interfere the corrent functioning of $A + B + C$, the system produces a mutated system, possibly of higher complexity than its parent, i.e.,

$$
\begin{aligned}
(A + B + C) &+ I_{A+B+C+F} \\
&\rightarrow (A + B + C) + I_{A+B+C+F} \\
&\quad + (A + B + C + F) + I_{A+B+C+F},
\end{aligned} \qquad (5)
$$

where F is an additional component that happened to emerge in the tape by mutation. The existence of and the relationship between phenotypes ($A+B+C$, $A+B+C+F$) and genotypes (I_{A+B+C}, $I_{A+B+C+F}$) illustrated in this formulation hold a close resemblance with those found in reproduction and variation of real (asexually reproducing) organisms.

Later, von Neumann tried to show that such a counterexample *does* exist, at least theoretically, by implementing his very complex universal constructor in cellular

automata—an artificial but totally mechanistic universe. Although he passed away before the completion of this work, it was posthumously completed and published by Burks in 1966 (von Neumann 1966).

Almost needless to say, von Neumann's idea of self-replicating automata was so profound that it was followed by a number of succeeding studies that ranges from cellular automata based self-replicating systems to core memory based evolutionary systems, which now forms one of the central parts of artificial life (Sipper 1998).

Universal Computation vs. Universal Construction

Many think of von Neumann's work on self-replicating machines as something different, and even weird, compared to his other (probably more famous) achievements in computer science, game theory, quantum physics, nuclear physics, and so on. However, it is worth pointing out that his idea of universal construction with instruction tapes has a fundamental correspondence with his another idea of universal computation with program codes stored in memory. These two are actually the same in that a system contains an arbitrary sequense of information inside itself, and provided it has a universal capability, it can immitate the behavior of any arbitrary system described in the sequence. Presumably von Neumann had no substantial distinction between these two in his mind when he considered these ideas.

There is a practical difference, however, between the universal computer and the universal constructor in terms of the requirement for the robustness of the systems to perturbations. The universal computer does not have to be robust by itself, because the computation theory assumes discrete mathematical entities and their state transition in a noiseless world. The responsibility of realizing such a noiseless condition is primarily on the device manufacturing side and not on the system itself. On the other hand, the universal constructor should be robust *per se*, because its construction capability best makes sense in our real, three-dimensional kinematic universe. Thus the system should be able to deal with more or less continuous physical entities that inevitably involve fructuations and uncertainty. In such settings, the system itself *is* responsible for the robustness of its workings.

The robust universal construction is apparently a very difficult problem to attain. Von Neumann himself tried to consider a kinematic model at first, but later he abandoned it and switched to cellular automata, following Ulam's suggestions, to avoid this difficulty. Although his work was still monumental and stimulative enough even on cellular automata, the lack of robustness was crucial when considered as a model of real biological and/or engineering systems. Therefore, in spite that his univer-

sal computer has been implemented hundreds of millions of times (i.e., computers we use nowadays), there has been *no* physical implementation of his universal constructor (Friedman 1996). This fact is now recognized as one of the grand challenges in artificial life (Bedau *et al.* 2000). So the relevant question to ask here is: *How can we enhance the robustness of self-replication processes?*

Workplace Construction

In this section I present a simple extension of von Neumann's theoretical model, which I tentatively call a *workplace construction model*, to illustrate a possible solution for the question mentioned above. The key idea is the introduction of an additional subsystem that constructs a "workplace" prior to automaton construction. This model can be described within von Neumann's framework by putting the added subsystem as a variant of automaton F in his formulation, except for three additional assumptions I am going to make in what follows.

The first assumption is a reasonable statement that we all know empirically:

Assumption 1: The construction and duplication processes that are originally sensitive to perturbations can be substantially stabilized by putting them on some solid supporting structure, or *workplace.*

For example, drawing a straight line is rather difficult by a free hand but is quite easy and precise by using a ruler. Using this assumption, I symbolically represent such stabilized processes by

$$A + I_X + S_X^A \quad \Rightarrow \quad A + I_X + S_X^A + X, \qquad (6)$$
$$B + I_X + S_X^B \quad \Rightarrow \quad B + I_X + S_X^B + I_X, \qquad (7)$$

$$(A + B + C) + I_X + S_X^A + S_X^B$$
$$\Rightarrow (A + B + C) + I_X + S_X^A + S_X^B + X + I_X, \ (8)$$

where S_X^A and S_X^B represent the workplaces that support automaton construction and tape duplication, respectively. Subscript X means that the size and/or shape of the workplaces may depend on the product, if not always. The bold right arrow "$\Rightarrow$" denotes that the process is significantly robust to perturbations, i.e., the outcome is not affected virtually by small but positive amount of perturbations. Here let us keep ourselves quite loose in evaluating the robustness; I just classify processes as either "robust" or "sensitive" to perturbations. Such a simplification enables us to think about the problem more clearly and concisely. Detailed quantitative analysis of the robustness is far beyond the range of this brief paper.

The above forms (6)–(8) look quite similar to the original ones (1)–(3), except for the addition of the two workplaces, so one may want to simply apply $X = (A + B + C) + S_X^A + S_X^B$ to the last form to obtain a

robust self-replication process. However, this would result in a non-trivial problem: By definition, workplaces are assumed to hold and support the construction process of the product, so its size in general should be equal to or greater than that of the product. Therefore, if one lets $X = (A + B + C) + S_X^A + S_X^B$, then, for example, S_X^A must be large enough to hold the entire product X that includes S_X^A, resulting in a vicious circle that never closes.

Note that this problem is deeply related to the complexity decreasing rule empirically seen in engineering: To obtain a highly organized product, one needs a stabilized and well controlled manufacturing process, which requires a machine that can control the local environment that should be equal to or larger than the product. Part of the reason why von Neumann's machine can seemingly construct a product more complex (larger) than itself is, in this context, because his machine is implicitly supported by an infinite array of cellular automata that virtually works as a solid workplace. In other words, he intended to overcome this empirical rule, but what he actually chose was avoiding it by putting away the crux of the issue out of consideration.

In order to resolve the above problem, here I make the second assumption:

Assumption 2: Workplaces are generally made of a simple but extensive repetition of the same kind of components, so they can be generated with enough preciseness by itself (without another workplace) even under perturbations.

Readers may wonder if making this assumption might be just another way of avoiding the problem. Since the main aim of this paper is to show the argument and promote discussions on it, I do not mean this is the one and only right solution. Nonetheless, many empirical observations seem to support this assumption. For example, one can create a straight line in a robust fashion by combining tiny rods into triangular meshes or trusses. As long as the tiny rods are at the same length, the outcome can be precise enough. This is intuitively because the product of this process is not complex; it can be produced by a repetition of the same simple tasks, which is not the case for more sophisticated processes like automaton construction.

With this second assumption, I introduce a set of new robust construction processes, i.e.,

$$R^A + I_X \quad \Rightarrow \quad R^A + I_X + S_X^A, \qquad (9)$$
$$R^B + I_X \quad \Rightarrow \quad R^B + I_X + S_X^B, \qquad (10)$$

where R^A and R^B are subsystems that estimate the size of needed workplaces for A and B from the description tape and construct them in a robust fashion. Note that the exact estimation of the size of needed workplaces

might be another complex task that needs another workplace to carry it on. We thus need the third assumption:

Assumption 3: The size (or the upper bound of the size) of workplaces for construction and duplication processes can be estimated from the instruction tape in a simple operation, even under perturbations.

This assumption is much less obvious than the previous two, and should be subject to discussion and verification. If it would not be the case in some condition, the arguments developed below would lose its universality, but I believe it would still have sufficient implications for how to create robust systems.

Provided the third assumption applies, the previous form of stabilized automaton construction (8) can be rewritten as

$$(A + B + R^A + R^B + C) + I_X$$
$$\Rightarrow (A + B + R^A + R^B + C) + I_X + S_X^A + S_X^B \qquad (11)$$
$$\Rightarrow (A + B + R^A + R^B + C) + I_X + S_X^A + S_X^B + X + I_X,$$

where the entire system $(A + B + R^A + R^B + C)$, which we call G hereafter, is now capable of both constructing X and duplicating I_X in a robust manner, while keeping its own size finite and independent of what is written in I_X. Note that the function of controller C is more complex than before; it now has to coordinate the behaviors of four other subsystems, A, B, R^A, and R^B.

Then, finally, we put the system G itself into X in the above form, to obtain

$$G + I_G \quad \Rightarrow \quad G + I_G + S_G^A + S_G^B$$
$$\Rightarrow \quad G + I_G + S_G^A + S_G^B + G + I_G. \quad (12)$$

This represents a robust self-replication process of G, which also produces some byproducts (S_G^A and S_G^B).

Discussion

The workplace construction model presented above is by no means theoretically proven. It uses three assumptions, each of which must be carefully checked about its validity. Whether such a robust automaton can be actually implemented is another key issue to be investigated, probably with a lot of efforts, just like what von Neumann did with cellular automata to show a concrete example of his self-replicating machine.

Despite these problems all, one can obtain a supportive implication for the model from a variety of phenomena observed in real biological systems at various scales. They imply the applicability and effectiveness of the central idea of the presented model. For example, at the smallest level, the formation of cell membranes is probably the most fundamental instance of workplace construction; it isolates the metabolic process of the cell

from environmental perturbations and keeps all the sensitive parts at the same place without diffusion. At a much higher level, niche construction seen in ecology (Laland, Odling-Smee & Feldman 2000) is another clear instance of workplace construction, such as dam building of beavers.

More directly relevant cases can be found in animals that are about to produce their next generation. For example, firm eggshells of birds and reptiles are probably the most direct example of workplace construction. Although the automaton constructors are embedded in the fetuses in these cases, the purpose of eggshells are exactly the same as is discussed in this paper, i.e., to stabilize the process of offspring construction by isolating it from the outside and holding it on a solid structure. More sophisticated instance is the uterus of mammals, where the workplace is included in the parent's body, but can be refurbished and extended as needed to hold its offspring under construction. It is also interesting that the production of S_G^A and S_G^B in (12) correctly captures the nature of the process that these workplaces constructed in real organisms (eggshells, endometriums, placentas, etc.) are all just for temporary use and they will eventually become garbages after birth of the offspring.

To conclude, the only goal of this short paper is to shed light on this trivial fact: the more complex a system becomes, the better controlled local environment the system needs in order to construct its replica under perturbations. Biological organisms seem to evolve such sophisticated "workplaces" for their survival and prosperity. However, this point has been long missing in the earlier studies of artificial self-replicating systems, and it must be taken into account to proceed toward the next step of artificial life.

References

Bedau, M. A., McCaskill, J. S., Packard, N. H., Rasmussen, S., Adami, C., Green, D. G., Ikegami, T., Kaneko, K., and Ray, T. S. 2000. Open Problems in Artificial Life. *Artificial Life* 6:363–376.

Friedman, G. 1996. The Space Studies Institute View on Self-Replication. *The Assembler: Newsletter of the Molecular Manufacturing Shortcut Group of the National Space Society* 4:4 (http://www.islandone.org/MMSG/9612.html).

Laland, K. N., Odling-Smee, J., and Feldman, M. W. 2000. Niche Construction, Biological Evolution, and Cultural Change. *Behav. Brain Sci.* 23:131–175.

McMullin, B. 2000. John von Neumann and the Evolutionary Growth of Complexity: Looking Backwards, Looking Forwards... In Bedau, M. A., McCaskill, J. S., Packard, N. H., and Rasmussen, S., eds., *Artificial Life VII: Proceedings of the Seventh International Conference on Artificial Life,* 467–476. Cambridge, MA: MIT Press.

Sipper, M. 1998. Fifty Years of Research on Self-Replication: An Overview. *Artificial Life* 4:237–257.

von Neumann, J. 1951. The General and Logical Theory of Automata. In Jeffress, L. A., ed., *Cerebral Mechanisms in Behavior—The Hixon Symposium,* 1–41. New York: John Wiley. Originally presented in September, 1948. Also collected in Aspray, W., and Burks, A. W., eds., *Papers of John von Neumann on Computing and Computer Theory,* 391–431. 1987. Cambridge, MA: MIT Press.

von Neumann, J. 1966. *Theory of Self-Reproducing Automata.* Urbana, IL: University of Illinois Press. Edited and completed by A. W. Burks.