Modelling Autopoiesis: Harder Than It May Seem!

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

Abstract

This article is an informal presentation of recent attempts to reproduce the phenomena of "computational autopoiesis" first investigated by Francisco Varela, Humberto Maturana and Ricardo Uribe almost 25 years ago (Varela, Maturana & Uribe 1974). A more formal discussion is available in (McMullin & Varela 1997).

Presented at ***Biology, Cognition, Language & Society: An International Symposium on Autopoiesis,*** *Belo Horizonte, Brazil, November 1997.*

Dublin City University
School of Electronic Engineering
Technical Report Number: `bmcm9704`

1 Prologue: The Scientific Enterprise

We might say that science in the modern sense was born when Galileo Galilei turned his first crude telescope on the night sky and discovered, among many other wonders, the moons of Jupiter. In one stroke a vast host of previous ideas about the universe (and our own place within it) became untenable. Or at least, so it seemed to Galileo. Unfortunately, there was one small problem: it also appeared that Galileo was the only one who could actually see his starry messenger...

So, it is perhaps better to say that modern science, or the scientific "method" *really* arose, not when Galileo observed the sky, not even when he published his results, but when he willingly—enthusiastically, even—accepted the challenge of *independent testing*. He accepted, in particular, that no one else should accept his claims at face value, or by virtue of some authority; but rather, that they should try for themselves. And when these independent tests initially failed, he made available more detailed information on his instruments, and even the instruments themselves. This allowed other influential individuals to discover for themselves that Galileo's wild claims had some basis in fact; and in time, it permitted improved design and understanding of telescopes, and of the critical conditions necessary to the various observational phenomena which Galileo had described.

To this day, openness to independent, critical, testing is a hallmark of what we call science. But even as this core principle has remained unchanged, scientific *practise* has had to continually debate and refine its interpretation, in the face of new technological apparatus, new theoretical understanding, and even—occasionally—the opening up of whole new domains of science. This article describes one recent installment in the debate. It's a real story of real scientists, and real science (in all its messy human imperfection)—somewhat in contrast to the sanitised and platonic science which scientists sometimes prefer to advertise. But it has a happy ending, at least, and maybe even a good old-fashioned moral...

2 ALife: Hype or Science?

"Artificial Life", or *ALife*, is a newly named, if not newly established, domain of intellectual enquiry. It has strong ties to the *Santa Fe Institute* through the original series of ALife Workshops organised at the Institute by resident faculty member Chris Langton (Langton 1989, Langton, Taylor, Farmer & Rasmussen 1992). There is a complementary European conference series, the first of which was jointly initiated by Francisco Varela and Paul Bourgine (Varela & Bourgine 1992). More than one "popular" book has been published on the subject.

At the very least, ALife qualifies as hype; but is there still "proper" science behind the hype?

The guiding inspiration of ALife is to realise "life-like" phenomena in artificial media (especially computers), and thus both to improve our understanding of natural living systems and to open up entirely new technological opportunities for artificial living systems.

ALife is then, in major part, a *synthetic* enterprise—the attempt to create artificial life. A typical report of research in the field may consist in the claim that a certain (life-like) phenomenon will be exhibited by a certain configuration or arrangement of some artificial system; or, more specifically, by the execution of a certain class of computer model.

Because of this synthetic aspect of the work, the nature of scientific testing or criticism can acquire a somewhat different flavour from more traditional science. It *is* sometimes the case that ALife models are directly related to particular natural living phenomena; and in those case, claims arising from ALife research may be testable by comparison with empirical biological data. However, there is also a class of ALife model in which the relationship to specific living systems is very remote or abstract. In this case, the objective is not so much to explain a particular natural phenomenon, as to claim that a certain kind of phenomenon will be robustly exhibited by a broad variety of systems having a certain abstract or formal structure. In this case, critical testing need not involve comparing with pre-existing empirical phenomena, but may rather require independent synthesis of artificial systems, conforming to the stated abstract structure, and testing whether they do, in fact, demonstrate the claimed phenomena.

It should be clear that, in principle at least, ALife can thus be just as "scientific" as its longer established ancestors; but equally, the practise and methodology of this synthetic kind of critical testing may need adjustment.

3 A Case Study: Computational Autopoiesis

3.1 The Original Model

In the early 1970's (well ahead of the modern rebirth of ALife) two Chilean biologists, Humberto Maturana and Francisco Varela, were struggling with the question of how a living system (such as a bacterium) is different from an arbitrary assemblage of its raw molecular constituents. More precisely, what is it that is so special about the *organisation* of the molecular constituents of a living system that gives rise to the living phenomenology?

Their tentative answer was that living systems are distinguished by a special class of organisation which they termed *autopoiesis*, from the Greek roots for "self-producing". Roughly speaking, an autopoietic system is a closed network of chemical reactions, such that every molecular constituent of the system is a product of one of the reactions in the network, and the system establishes and maintains a spatial boundary which serves to concentrate the reactants and sustain the network.

This concept is still somewhat vague and it is not clear just how much of the living phenomenology it can illuminate; but it is certainly a creative and original approach to this very difficult problem.

In any case, in an effort to make their ideas more concrete, Varela and Maturana also developed a "minimal model" for the kinds of components and interactions that seem to be required to support autopoietic organisation; that is to say, they described an abstract "artificial chemistry" in which autopoietic systems could both spontaneously emerge and sustain themselves. Finally, with another colleague, Ricardo Uribe, they developed a specific computer model or realisation of this artificial chemistry, and demonstrated that it did indeed support at least the basic autopoietic phenomenology.

Clearly this work falls under the general class of synthetic ALife work described earlier: there is a claim that certain life-like phenomena (autopoietic organisation) can arise in any system which realises a certain set of abstract "chemical" interactions. And equally, although this particular claim cannot be meaningfully tested against natural, biological, data, it is still perfectly scientific in the sense that it can be tested by the synthesis of independent realisations (computer models) of the same abstract chemistry.

Of course, all that remained was for someone to actually do such tests...

3.2 The Problems

I personally became interested in the computer modelling of autopoiesis about five years ago (McMullin 1994). While the phenomenology described for the basic system was quite rudimentary I hoped that it might be possible to extend it to support systems which can grow and reproduce via a crude fissioning mechanism. This could allow an approach to modelling and understanding the evolutionary emergence of genetic reproductive architecture—and thus the emergence of "evolvability".

However: the first step would clearly have to be to reproduce (sic!) the basic phenomena of the original model.

At first sight, this should not have been difficult. The originally published description of the work included both an informal description of the essential interactions in the artificial chemistry, and a more detailed algorithmic description of their particular computer program which realised this chemistry. In the noble tradition of the professional academic researcher, I therefore simply presented this paper to an unfortunate graduate student (Hyder Aswad) and asked him to re-implement it.

Which he did.

Unfortunately, this re-implementation completely failed to exhibit the original phenomena.

I must confess, to my shame, that (remaining true to the aforementioned noble academic tradition) my immediate reaction was to blame my poor graduate student, and declare that he must have made some mistake (probably more than one!) in his re-implementation. He, on the other hand, rather trenchantly insisted that he had checked his work with great care. So, to prove my point (and, as an added benefit, to underscore why I was the exalted professor, and he the humble graduate student) I duly rolled up by virtual shirt sleeves and proceeded to take his program apart, line by line, to identify exactly where he had gone wrong.

Unfortunately (for my self-respect if not the progress of science!) it was now my turn to fail. Miserably. I could find absolutely nothing wrong. Granted, some aspects of the original description were unclear or ambiguous; some details of the implementation might have been done more efficiently. But none of these things could apparently explain the failure of the model.

At this point, I contacted Francisco Varela, whom I had met briefly before. He was able to tell me that, regrettably, the original program had been lost; but he directed me to a student of his who had recently re-implemented the model as an exercise. The student, Francois Jullien, kindly forwarded his Pascal program code to me—noting, however, that it had been developed on an obsolete computer platform that he no longer had available, so he could not run the program himself anymore! Strangely enough, I did not have access to this obsolete platform either, but I hoped to glean some insight from examination of his code. Granted, the identifiers and comments in the code where all in French, with which language, unfortunately, I have only very modest familiarity; but Francois very generously agreed to provide an overall technical description of the program in English.

In parallel with this I also sought other researchers who might have successfully reproduced the original phenomena. A literature search revealed only one substantive example, by Milan Zeleny (Zeleny 1977, Zelany & Pierre 1976). But at least the code from this example had apparently been made publically available. Admittedly, it was written in a somewhat obscure and specialist computer language (APL), but I was prepared to deal with that if necessary. However, on contacting Dr. Zeleny, it transpired that, although his code *had* been made freely available at the time, this was quite some time in the past ... and all copies had now, as far as he knew, been lost.

Broadcasting a message on a related Internet mailing list elicited one additional lead: John Mingers, a researcher based in the UK, had also independently re-implemented this model, as an exercise. But, as with Jullien's version, this had been developed on a computer platform that was now obsolete and no longer available. In fact, only a hardcopy of the code, rather than a machine readable version had been retained. However, Mingers had recently had the program manually rekeyed from the hardcopy, though had not made any attempt to get it running. He was willing to give me a copy of this, and, indeed, collaborate with efforts to get it running again.

I studied the code from these two independent re-implementations carefully, but could identify no significant difference from that of my own student. It became clear that I would need to see either or both of these actually working if I was to make further progress. Over a period of time, and with assistance from both Jullien and Mingers, I ported both of their programs to a common platform which I had available, and got them "working" (Jullien & McMullin 1995, Mingers & McMullin 1997). That is, they executed, apparently as their respective authors had intended; but they did *not* exhibit the autopoietic phenomena described in the original publication. In fact, they behaved in essentially the same flawed manner as the program written by own student, Hyder Aswad!

At this point, frustration, and other demands on my time, meant that I had to put this project aside. Hyder Aswad also moved on to other things. However, I was conscious that these results were deeply unsatisfactory, and certainly intended to return to the project when an opportunity should arise.

After an hiatus of around 18 months, that opportunity did eventually arise when I was enabled to spend an extended sabbatical visit at the Santa Fe Institute (I am grateful especially to Chris Langton and John Holland for facilitating this).

My first step was to build yet another model of the artificial chemistry, this time layered on top of the general purpose simulation system called Swarm, recently developed at SFI, and programmed in Objective-C. This adopted a completely different (agent-based) simulation architecture, ignoring the original algorithm, and attempting to capture only the original qualitative chemistry. Not surprisingly, this again made no substantive difference: the original phenomena could not be reproduced with this model either.

At this point, the failures could be clearly related to a single core problem: a certain reaction of the artificial chemistry ("bonding"), when implemented according to the original specification, was taking effect "prematurely" relative to the putative autopoietic network.[1] It seemed that there must be something *missing* from the model; something that was not included in the published description, but that *was* present in the original simulation code. But how could this possibly be identified now, a quarter of a century after the original experiments had been carried out?

4 The Resolution

By a wonderful and utterly unpredictable chance, just as I contacted Francisco Varela again, to pass on my assessment that there was some crucial lacuna in the original published descriptions of this model, he received a shipment of papers from Santiago, dating from the general period when this work was carried out, and which he had thought to have been completely lost. Searching through these materials, he found a single hardcopy printout of a single version of the original simulation program, together with a short textual description of the model, which he believed to be at least roughly contemporary with the

[1] See the example in the appendix.

program listing! These materials are now permanently archived in (McMullin 1997*b*).

In context, of course, this was the intellectual equivalent of striking gold. If there was ever to be any possibility of definitively resolving what was really going on in the original program, then access to that program's code was the absolute *sine qua non*; and here, against all rational expectations, was, precisely, a snapshot of that program code. Granted, a rather grainy snapshot, in black and white; but vastly more than I had expected to have available to me!

Notwithstanding all that, I hesitated long and hard before committing to a detailed examination of this code. It was written in the long obsolete language, FORTRAN-IV. It was only sparsely commented, and such comments as there were, were in Spanish. Variable names was generally less than four characters long and had little obvious mnemonic content. The program was structured as a single monolithic block of code, rather than using procedural decomposition (i.e. subroutines). There was no guarantee that this particular version of the program even worked. Nor could I assess its likely relationship to the version actually used to generate the published results. Trying to get it working would require manual re-keying, which is intrinsically error prone. H owever: while all these were legitimate causes for concern, I had really exhausted all other avenues of attack, so that it seemed I had little to lose by attempting to decipher this program.

My strategy was to first rekey the program and attempt to get it "working" in some minimal sense, without making any serious effort to understand its detailed functioning; only if I could achieve this, and if the program exhibited the crucial phenomenon (i.e. that premature bonding was somehow avoided) would I invest the effort to analyse the program's mechanisms in detail.

Somewhat to my surprise, this strategy worked amazingly well. Only a very small number of keying errors occurred, most of which were identifiable from compiler diagnostics. There was only one "external" fragment of code missing, and its function could be reasonably inferred from the code to hand (it was a pseudo random number generator) so that it was possible to design a functional replacement. This did result in a program that could be executed; and, against my more pessimistic expectations, it *did* reproduce a variety of characteristic phenomena of the original published results (encouraging me to believe that this was, at the very least, a closely related version); and, most crucially of all, it exhibited the avoidance of premature bonding that had plagued all my attempts at reproducing the full-blown autopoietic phenomena.

Now knowing that the answer to my question was surely concealed *somewhere* in this program, I threw myself into a sustained attempt to understand it. A key factor here was that it was not necessary to grasp the program in its entirity, but only the implementation of one particular interaction. Thus a relatively cursory examination of the program quickly allowed me to narrow my search down to one relatively small section—and to identify exactly what was built into the program that was preventing the premature bonding. This turned out to be a special inhibition effect that only came into play in certain special—but critical—circumstances. This effect had not been mentioned in any of the published descriptions of the model, but seemed to be responsible for the key difference between the original program and all my attempts are reproducing it.

So—finally—I was ready to make the critical, *scientific*, test: I incorporated the appropriate analog of this new effect into my own, otherwise entirely independent, implementation of the model. I had to experimentally adjust other parameters of the model to take account of this, but in rather less than half an hour (and after only five years of trying!) I finally saw for myself, the full blown autopoietic phenomena reported in the original 1974 paper![2]

5 Epilogue

I think there are a few worthwhile points to this story.

First: that science, in its necessary sociality, relies on scientists being open and willing to engage in critical assessments of their own work. In this respect, I am left with an abiding admiration for Francisco Varela, who happily and enthusiastically took up the critical challenge with which I presented him, and was unfailingly courteous and open in his responses. The tale could never have had a satisfying resolution otherwise.

Secondly, that science is actually an art!

By which I mean that it is not a specific method or process or algorithm by which knowledge is somehow guaranteed, but is rather a challenge and an adventure, where failure is at least as common as success, and luck, intuition, and entirely informal judgments play a pervasive role. I imagine that most scientists already know all this; but we somehow feel

[2] Some illustrations of the rediscovered phenomena are shown in the appendix.

a need to disguise or hide it. Perhaps it's time to come clean.

Finally, that science is *open and unlimited.* Entirely new domains of enquiry, such as ALife, can and do emerge, and can be subjected to the hard scientific criterion of independent, critical, testing.

References

Jullien, F. & McMullin, B. (1995), 'FRJ's Simple Autopoiesis Program'. Program source in Pascal, for MS-DOS platform. `ftp://ftp.eeng.dcu.ie/pub/alife/frj_sap`

Langton, C. G., ed. (1989), *Artifical Life*, Vol. VI of *Series: Sante Fe Institute Studies in the Sciences of Complexity*, Addison-Wesley Publishing Company, Inc., Redwood City, California. Proceedings of an interdisciplinary workshop on the synthesis and simulation of living systems held September, 1987, in Los Alamos, New Mexico.

Langton, C. G., Taylor, C., Farmer, J. D. & Rasmussen, S., eds (1992), *Artifical Life II*, Vol. X of *Series: Sante Fe Institute Studies in the Sciences of Complexity*, Addison-Wesley Publishing Company, Inc., Redwood City, California. Proceedings of the workshop on Artificial Life held February, 1990, in Sante Fe, New Mexico.

McMullin, B. (1994), Artificial Darwinism: The Very Idea!, *in* B. McMullin & N. Murphy, eds, 'Autopoiesis & Perception', Dublin City University, Dublin, pp. 71–94. Proceedings of a workshop held in Dublin City University, August 25th & 26th 1992. Report `bmcm9401`. `ftp://ftp.eeng.dcu.ie/pub/alife /bmcm9401/mcmullin.ps.Z`

McMullin, B. (1997*a*), 'Code McMullin: The Case of the Independent Test', *Bulletin of the Santa Fe Institute* **12**(2). Summer 1997. `http://www.santafe.edu/sfi/publications/ Bulletins/bulletin-summer97/code.html`

McMullin, B. (1997*b*), Computational Autopoiesis: The Original Algorithm, Working Paper `97-01-001`, Santa Fe Institute, Santa Fe, NM 87501, USA. `http://www.santafe.edu/sfi/publications/ Working-Papers/97-01-001/`

McMullin, B. & Varela, F. J. (1997), Rediscovering Computational Autopoiesis, *in* P. Husbands & I. Harvey, eds, 'Proceedings of the Fourth European Conference on Artificial Life', MIT Press. `http://www.eeng.dcu.ie/~alife/bmcm-ecal97`

Mingers, J. & McMullin, B. (1997), 'JM's Simple Autopoiesis Program'. Program source in Pascal, for MS-DOS platform. `ftp://ftp.eeng.dcu.ie/pub/alife/jm_sap`

Varela, F. J. & Bourgine, P., eds (1992), *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life*, Series: Complex Adaptive Systems, MIT Press, Cambridge.

Varela, F. J., Maturana, H. R. & Uribe, R. (1974), 'Autopoiesis: The Organization of Living Systems, its Characterization and a Model', *BioSystems* **5**, 187–196.

Zelany, M. & Pierre, N. A. (1976), Simulation of Self-Renewing Systems, *in* E. Jantsch & C. H. Waddington, eds, 'Evolution and Consciousness: Human Systems in Transition', Addison-Wesley Publishing Company, Reading, Massachusetts, chapter 7, pp. 150–165.

Zeleny, M. (1977), 'Self-Organization of Living Systems: A Formal Model of Autopoiesis', *International Journal of General Systems* **4**, 13–28.

Acknowledgments

Initial work on re-implementing a computational model of autopoiesis was carried out by my former Hyder Aswad. I am grateful to Francoise Jullien and John Mingers for making available their independent re-implementations of the model of (Varela et al. 1974). I especially acknowledge the open and generous assistance of Francisco Varela.

This work described here was carried out while I was a visiting researcher (1996/97) with the Swarm Project Group at the Santa Fe Institute. I am very grateful to Chris Langton of SFI for hosting that visit. The visit was made possible by generous support from Dublin City University and the Swarm Project. Swarm has been supported by Grant No. N00014-94-1-G014 from the Department of the Navy, Naval Research Laboratory, acting in cooperation with the Advanced Research Projects Agency. Swarm has also benefited from earlier support from The Carol O'Donnell Foundation and from Mr. and Mrs. Michael Grantham.

Finally, I thank the organisers and sponsors of **Biology, Cognition, Language & Society: An International Symposium on Autopoiesis,** Belo

Horizonte, Brazil, November 1997, for so kindly inviting me to participate in this symposium.

All errors in the current article remain, of course, my own responsibility.

Revision History

This is an edited version of an article originally published in the *Bulletin* of the Santa Fe Institute (McMullin 1997*a*).

Retrieval

The resources comprising this article are electronically retrievable, in various formats, via the World Wide Web, from:

`http://www.eeng.dcu.ie/~alife/bmcm9704/`

Copyright

Appendix: Model Behaviour

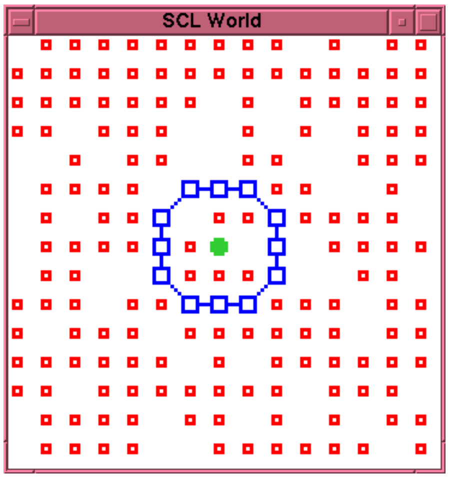

Figure 1: Artificial Cell: Time 0

The Artificial Cell: Substrate particles (red) can diffuse through the membrane (green), where they react under the influence of the catalyst particle to produce more membrane particles. These are trapped by the membrane and thus build up in concentration, held ready in reserve to effect a repair to the membrane whenever it should rupture. Such repair re-establishes the conditions for the continued maintenance of the cell.

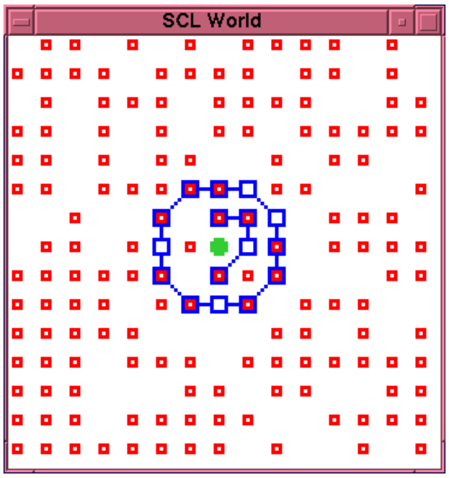

Figure 2: Experiment 1: Time 30

Premature Bonding: However, with the model as *originally* described, the new membrane fragments spontaneously and prematurely bond together, *inside* the cell. This makes them immobile and thus unavailable to migrate to a rupture site and effect a repair.

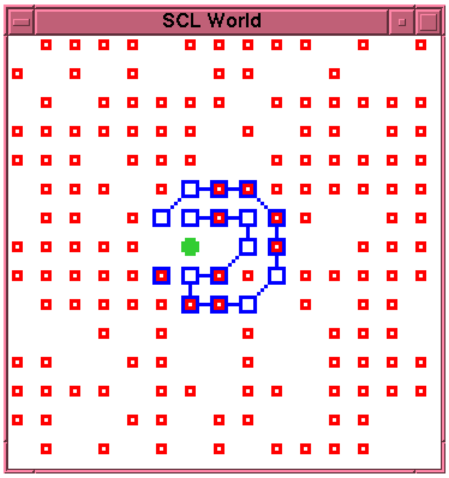

Figure 3: Experiment 1: Time 238

Degeneration: When the cell ruptures, instead of a repair, the interior fragment gets incorporated, leaving the catalyst particle free. The conditions for continued metabolism no longer exist: the cell is effectively dead.

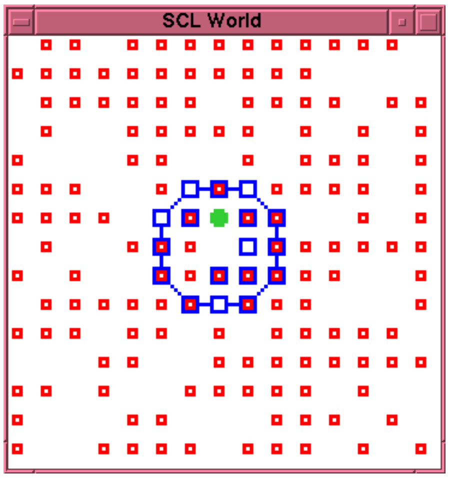

Figure 4: Experiment 2: Time 170

Bond Inhibition: The newly rediscovered effect stipulates that free membrane particles may not bond if they are immediately adjacent to already doubly bonded membrane particles. This prevents premature bonding and permits the build up of a reserve of free, mobile, membrane particles.

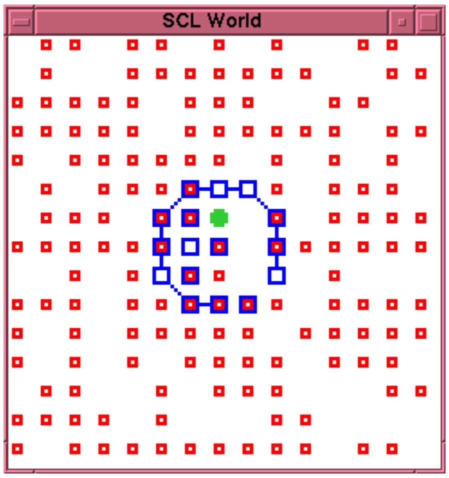

Figure 5: Experiment 2: Time 217

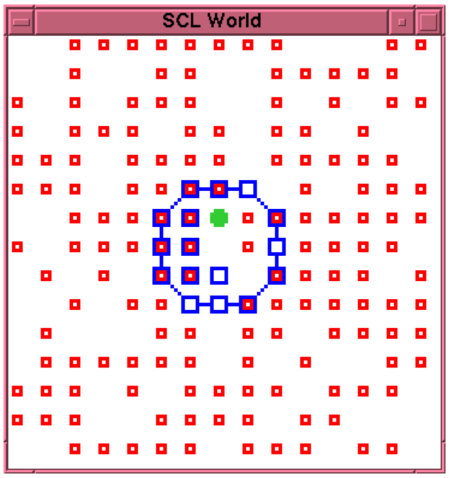

Figure 6: Experiment 2: Time 226

Self-Repair: At time 180 the mebrane ruptures; by time 217 a replacement particle drifts into position, and at time 226 the repair is successfully completed.