

Reality is Stranger than Fiction

What can Artificial Life do about Advances in Biology?

Luis M. Rocha

School of Informatics and Cognitive Science Program, Indiana University, USA

and

FLAD Computational Biology Collaboratorium, Instituto Gulbenkian de Ciencia, Portugal

Prepared for and presented in the “Biocomplexity” discussion section at the 9th *European Conference on Artificial Life*, September 12, 2007 in Lisbon, Portugal

“By extending the empirical foundation upon which biology is based *beyond* the carbon-chain life that has evolved on Earth, Artificial Life can contribute to theoretical biology by locating *life-as-we-know-it* within the larger picture of *life-as-it-could-be*”. [Langton, 1998, page 1]

From Langton’s original artificial life manifesto, the field was largely expected to free us from the confines of “life-as-we-know-it” and its specific biochemistry. The idea of “life-as-it-could-be” gave us a scientific methodology to consider and study the general principles of life at large. The main assumption of the field was that instead of focusing on the carbon-based, living organization, life could be better explained by synthesizing its “logical forms” from simple machines [Langton, 1989, page 11]—where, “fictional” machines substituted real biochemistry. The expectation was that this “out-of-the-box”, synthetic methodology would gain us a wider scientific understanding of life. We would be able to entertain alternative scenarios for life, challenge the dogmas of biology, and ultimately discover the design principles of life.

Interestingly, during the 20 years since the first artificial life workshop, biology witnessed tremendous advances in our understanding of life. True, biology operates at a completely different scale of funding and in a much larger community base than artificial life (the impact factors of key journals in both fields differ by an order of magnitude). But, still, it is from biology, not artificial life, that the strangest and most exciting discoveries and design principles of life arise today. Consider looking at the last number of *Nature*, with the quite apropos editorial title “Life as We Know it” [Vol. 449, 1], arguing for a comparative genomics approach, with articles, for instance, advancing towards evolutionary principles of gene duplication [Wapinski et al, 2007]. Publications in the last week in *PLoS. Biol.*, also presented new evidence towards updating or discovering general principles of life: for instance, Venter’s sequencing of his diploid genome, which updates our expectations of differences in chromosome pairs [Levy et al, 2007]), and the Ahituv et al [2007] study that challenges the idea that ultraconserved DNA (across species) must be functional.

It is good to notice that this sort of work is not so much an exception, but has been a signature of research in the biosciences in the last decade or more. Consider cases such as the discovery of DNA transfer from bacteria to the fly [Dunning Hotopp, 2007], extra-genomic inheritance in *Arabidopsis* [Lolle et al, 2005], or the profound importance of non-coding RNA in life which is a major player in, among other features, patterning [Martello et al, 2007], essential gene regulation [Mattick, 2005], development [Mattick, 2007], epigenetic neural development and modulation [Mehler and Mattick, 2007], eukariotic complexity [Taft et al, 2007], etc. Moreover, advances such as these do not seem to be mere epiphenomena of a specific life form. Indeed, they point at important organization principles—as those that artificial life was supposed to provide. When we discover that non-transcribed RNA is involved in extra-genomic inheritance or that most of the evolutionary innovation responsible for differences between marsupials and placental mammals occurs in non-protein coding DNA [Mikkelsen et al, 2007], some fundamental principles of the living organization are to be re-thought: the generalized genotype-phenotype mappings on which most of artificial life is based on, are just not enough to capture the principles of life as we know it.

One could go on and on about many other advances in biology. We can also point to themes at the forefront of (bio)complexity theory that go largely overlooked in artificial life—though not completely (i.e. [Calabretta et al, 2000; Hintze & Adami, 2007]). Perhaps the key topic in complexity theory today is that of modularity in evolution [Schlosser & Wagner, 2004] and in networks [Newman, 2006; Guimerà et al 2007]. Nonetheless, looking at the papers accepted for the main sections of the latest Alife and ECAL conferences, it is easy to see that most papers, not only do not discover or even address such issues, but largely trade in biological and computational concepts that have not changed much since the field's inception (see list of top themes and terms in appendix). Is artificial life trapped in the (evolutionary) biology of twenty years ago? Why is reality stranger and more surprising than fiction?

Clearly, there has been very widely successful artificial life research. First and foremost, artificial life has been most successful as a means to study animal behavior, learning and cognition. Certainly, evolutionary robotics and embodied cognition have had an impact in cognitive science. But is artificial life simply a better way to do artificial intelligence? Moreover, one could argue that given the embodied nature of evolutionary robotics, it would seem that it is bound to some kind of material reality, rather than synthesized by constituent “logical forms” as Langton initially suggested.

But what to do about the organization of life itself? Surely the idea of explaining the living organization was behind the origin of the field. For the purposes of this discussion, we must question ourselves why artificial life does not produce more and surprising results about the living organization? Certainly, there is sound research in the field with impact outside of it [e.g. Adami, 2006; Hintze & Adami, 2007]. But even the most successful research in artificial life rarely goes

beyond showing that artificial organisms can observe the same behaviors as their real counterparts (i.e. selective pressures, epistasis, etc.). A problem for the field is that as biotechnology gains more and more control of cellular processes, it is reasonable to ask what can one do with artificial organisms that one cannot do with real bacteria? For instance, recent studies of the evolutionary speed towards beneficial mutations were quite effectively done with E-coli [Perfeito et al, 2007], pointing to a much larger rate of beneficial mutations in bacteria than previously thought, and shedding new light on the general principal of clonal interference.

The point of this short statement is to discuss at this conference [ECAL 2007], how biocomplexity is dealt with in artificial life, twenty years after the field's inception. Certainly the community can think of a variety of responses to this lack of new principles of life coming out of research in artificial life—even in theoretical biology. One concept that I venture may need updating in artificial life is its view of the genotype/phenotype relationship. Langton proposed that we generalize this relationship, but this meant that research in the field largely regarded the two as indistinguishable. While this move at fist glance seems appropriate to deal with the complexity of genomic-proteomic interaction, it prevents us from studying the specific roles each plays in the living organization. Genotype and phenotype are intertwined in a complex manner, but each operates under different principles that are often overlooked in artificial life. Thus, artificial life rarely approaches issues of genomic structure and regulation, or the co-existence of DNA and RNA as different types of informational carriers. This could well be because artificial life models seem to trade most often on the concept of Mendelian gene than on the molecular biology gene. In other words, artificial life models tend to regard genes solely as mechanisms of generational (vertical) inheritance, rather than as (informational) mechanisms of ontogenetic (horizontal) development, regulation, maintenance, phenotypic plasticity, and response to environmental change. This way, most artificial life models do not test, or even deal with, possible genomic structure architectures and their impact on development and evolution. This is a big shortcoming in the field since, as we have seen in the last two decades, the molecular biology gene and the genomic structure it implies are behind many essential principles of life—from hypersomatic mutation in vertebrate immunity to speciation.

Additionally, it is most often the case that artificial organisms in artificial life models are designed with many top-down features, rather than emerging out of artificial biochemical machines. For instance, typically the genes of artificial organisms encode pre-defined computer operations. Not only is the encoding pre-defined, but the function of individual genes is also pre-programmed, rather than emergent from some artificial chemistry—what is typically emergent is the behavior of a collection of such “atomic” genes and genotypes.

It is interesting to note that when biologists were looking for the location of genetic information for inheritance, they naturally assumed that it would reside in proteins. They knew of DNA chemically, but its sheer inertness deemed it unfit for the job. It took some time (from Griffith's experiment in 1928 to Avery's in 1944) to realize that relative inertness was really the point. The biochemical

difference between highly inert memory molecules and highly reactive, functional ones, while often overlooked in artificial life as a design principle, is ultimately the hallmark of life. Perhaps, Langton's view of artificial life being built-up from simple machines, may have clouded the fact that life as we know it is made of biochemical constituents with very different roles: chiefly, DNA, RNA and proteins. Perhaps more attention should be directed to the "logical forms" of the lower level, structural constituents that produce life, before we can tackle "life-as-it-could-be" ?

ACKNOWLEDGMENTS

I am extremely grateful to Karola Stotz at Indiana University for exciting conversations that directly lead to this text. I also thank Janet Wiles for performing the *Leximancer* analysis, and Peter Todd for patiently listening as some of the ideas here discussed were germinating.

REFERENCES

- Ahituv, N., et al. [2007]. "Deletion of Ultraconserved Elements Yields Viable Mice". *PLoS Biology* Vol. 5, No. 9, e234.
- Calabretta, R., Nolfi, S., Parisi, D. and Wagner, G.P. [2000]. "Duplication of modules facilitates the evolution of functional specialization". *Artificial Life*, 6 , 69-84.
- Hintze, A. and C. Adami [2007]. "Evolution of complex modular biological networks". arXiv.org:0705.4674.
- Dunning Hotopp, J. C., et al. [2007] "Widespread Lateral Gene Transfer from Intracellular Bacteria to Multicellular Eukaryotes," *Science* doi:10.1126.
- Guimerà, R, Sales-Pardo, M, Amaral, LAN. [2007] "Classes of complex networks defined by role-to-role connectivity profiles". *Nature Phys.* 3, 63-69.
- Levy, S. et al. [2007] *PLoS Biol.* 5, e254.
- Lolle, S. J., et al [2005]. "Genome-wide non-mendelian inheritance of extra-genomic information in Arabidopsis", *Nature* 434, 505-509.
- Mikkelsen, T.S. et al [2007]. "Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences". *Nature* 447, 167-177. ,
- Martello, G. Et al [2007]. "MicroRNA control of Nodal signalling" *Nature*. doi:10.1038/nature06100;

Mattick, J.S. [2005] "The functional genomics of noncoding RNA". *Science*. 309 (5740): 1570-3.

Mattick JS.[2007]. "A new paradigm for developmental biology". *J Exp Biol.* ;210(Pt 9):1526-47

Mehler MF, Mattick JS. [2007]. "Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease". *Physiol Rev.* 87(3):799-823

Newman, M.E.J. [2006]. "Modularity and community structure in networks". *PNAS*. 103: 8577.

Perfeito L, L. Fernandes, C. Mota and I. Gordo [2007]. "Adaptive Mutations in Bacteria: High Rate and Small Effects". *Science* (in press)

Schlosser, G. and G. P. Wagner (Eds.) [2004]. *Modularity in Development and Evolution*.

Taft RJ, Pheasant M, Mattick JS. [2007]. "The relationship between non-protein-coding DNA and eukaryotic complexity". *Bioessays*. 29(3):288-99.

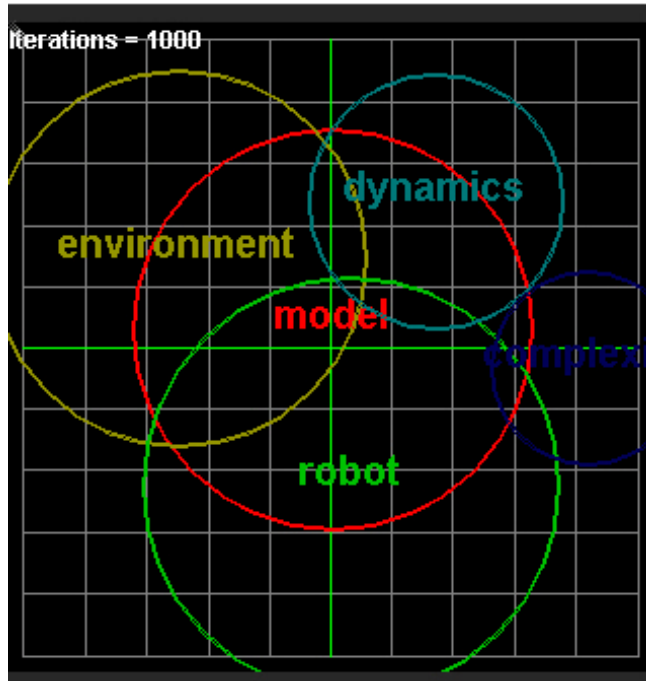
Wapinski, I., A. Pfeffer, N. Friedman, and A. Regev [2007]. "Natural history and evolutionary principles of gene duplication in fungi". *Nature* 449, 54-61.1.

APPENDIX:

Top themes extracted from all abstracts accepted to ECAL 2007, produced the Leximancer (courtesy of Janet Wiles)

Concept	Absolute Count	Relative Count
<u>model</u>	119	100%
<u>system</u>	95	79.8%
<u>evolution</u>	92	77.3%
<u>results</u>	64	53.7%
<u>environment</u>	58	48.7%
<u>behavior</u>	57	47.8%
<u>present</u>	57	47.8%
<u>network</u>	56	47%
<u>robot</u>	55	46.2%
<u>agents</u>	52	43.6%
<u>simulation</u>	50	42%
<u>process</u>	49	41.1%
<u>simple</u>	46	38.6%
<u>show</u>	43	36.1%
<u>mechanism</u>	41	34.4%
<u>complex</u>	41	34.4%
<u>dynamics</u>	39	32.7%
<u>artificial</u>	39	32.7%
<u>problem</u>	39	32.7%
<u>based</u>	36	30.2%
<u>learning</u>	36	30.2%
<u>approach</u>	33	27.7%
<u>population</u>	31	26%
<u>study</u>	31	26%
<u>genetic</u>	30	25.2%
<u>individual</u>	30	25.2%
<u>neural</u>	28	23.5%
<u>selection</u>	28	23.5%
<u>organisms</u>	27	22.6%
<u>method</u>	26	21.8%
<u>conditions</u>	26	21.8%
<u>level</u>	26	21.8%
<u>information</u>	25	21%
<u>social</u>	25	21%

Top Themes produced from Leximancer set at 65% coverage themes (courtesy of Janet Wiles)



Top co-occurring (stemmed) word pairs in abstracts

- neural--network
- chang--environ
- artifici--life
- simul--result
- autonom--robot
- evolutionari--algorithm
- evolutionari--robot
- comput--simul
- genet--algorithm
- robot--mobil
- cellular--automata
- interact--between
- artifici--chemistri
- agent--adapt
- pressur--select
- neural--control