In Proceedings of the 1999 Genetic and Evolutionary Computation Conference Workshop Program, Orlando, Florida, July 13, 1999, Annie S. Wu, ed., pp. 20-23.

Can Unrealistic Computer Models Illuminate Theoretical Biology?

Mark A. Bedau
Department of Philosophy
Reed College
3203 SE Woodstock Blvd, Portland OR 97202
mab@reed.edu
http://www.reed.edu/~mab
(503) 771-1112, ext. 7337

1 Deep Questions in Biology

Questions about the important essential properties of biological systems are both difficult to answer and worthwhile to try to answer. Here are three examples of deep open questions in theoretical biology:

1. Is robust multi-level emergent activity an intrinsic property of certain homeostatic self-organizing systems like cells or organisms, and if so, how is this possible?
2. Is open-ended adaptive evolution an intrinsic property of certain evolving systems like the biosphere, and if so, how?
3. Is unbounded complexity or diversity growth an intrinsic property of certain evolving systems like the biosphere, and if so, how?

These questions concern apparent fundamental properties of living systems—properties which, furthermore, seem to be shared by many other complex adaptive systems, such as the global economy. It is especially hard to address these questions, largely because they concern the global emergent behavior of overwhelmingly complex systems. One way to pursue answers is with a certain sort of *unrealistic* computational model. Although this may sound paradoxical, I shall argue that, properly understood, it makes perfect sense.

I could not agree more when Levin et al. [10] say that "[i]maginative and efficient computational approaches are essential in dealing with the overwhelming complexity of biological systems" (p. 341). There are at least two quite different kinds of computational models of complex biological systems. One strives for maximal fidelity to the details of particular natural systems, exploiting prodigious computer power to push the envelope on micro-mechanical realism. But I am interested in models which intentionally *abstract away* from the micro-details in real systems, models which are as *unrealistic* as possible. These models are common in the field of artificial life, and I discuss them here to balance the (appropriate) attention given to realistic computational models in theoretical biology.

2 Why Unrealistic Models

The models I have in mind are unrealistic in the sense that they abstract away from as many micro-level details as possible. The goal of this abstraction is to find the minimal set of properties sufficient to generate and explain the phenomena under investigation. So, for example, to illuminate multi-level emergence or unbounded adaptive creativity or unbounded complexity growth, one would seek a maximally abstract agent-based model in which these macro-scale phenomena emerge from the aggregate behavior of the local interactions of the micro-scale individuals. The more abstract the model, the simpler the explanation it provides. Also, the more abstract the model, the more broadly it applies, so the more unified the explanation it provides. Simplicity and unity are paramount virtues of explanations in any context, and unrealistic models produce explanatory simplicity and unity by drastic abstraction.

It is worth noting that the traditional mathematical models of theoretical biology are even more unrealistic because their differential equations abstract away from even the micro-level details present in agent-based artificial life models. For example, the Lotka-Volterra model of predation makes no commitment to the details of any particular predator or prey and ignores all factors like spatial structure, competitive interference between different predators, and the dangers prey face when searching for mates. The simplicity and universality enjoyed by traditional mathematical models exactly parallels those enjoyed by maximally abstract agent-based models. What forces the shift from mathematical models to agent-based models is the forbidding complexity of the systems under investigation and the consequent need to study their global behavior as emergent phenomena in computer simulations. Traditional mathematical models are *too* abstract to answer many questions about complex adaptive systems.

By the same token, too much abstraction can also cripple an agent-based model by squelching the appropriate behavior. For example, it might turn out that a non-trivial genotype-phenotype mapping is an essential property of any evolutionary model capable of producing open-ended adaptive evolution. If so, then any agent-based model that abstracts away from such mappings will be inadequate to explain open-ended evolution. Determining how much abstraction is too much is typically an empirical matter, to be settled by trial and error.

An unrealistic model can explain a certain type of behavior successfully only if it actually produces the desired behavior, of course. So, if the goal is to explain the behavior exhibited by some real-world system, the model's behavior will be realistic; what will be unrealistic is the model's simplified micro-structure. But the micro-structure of such unrealistic models will not be *completely* unrealistic. Although vastly simplified, a model can explain the behavior of a real-world system only if the model's micro-structure captures the *abstract form* of the target system's micro-structure. The micro-structure of unrealistic models *is* realistic when viewed at an abstract enough level.

Unrealistic agent-based computational models share a number of important virtues with more realistic computational models, such as the explicitness and precision forced by the computational methodology. But perhaps the most important virtues are these three:

- **Feasible.** Explicit mechanistic feasibility is a valuable discipline enforced by any computational methodology, preventing any explicit or implicit appeal to "magic" in the model. In other words, computationally feasibility is insurance against violations of naturalism. If a model is unfeasible, then its behavior will illuminate nothing since you will not be able to observe its behavior

- **Emergent.** The macro-scale phenomena under investigation emerge from the micro-scale details in the model. The micro-scale mechanisms in the model *produce* the macro-scale phenomena, and the model shows how this emergent phenomena responds to changes in fundamental parameters. For example, experimentation with Packard's Bugs discloses how evolutionary activity depends on the mutation rate [16]. The emergent nature of what these models model is a key reason why these models must be computational. There is no way to observe the macro-scale behavior implicit in the model other than making the behavior explicit through simulations.

- **Experimental.** The usual way to determine the generic properties of these models is experimental data collection, often parameter sweeps. It is particularly interesting to measure properties of the models that can meaningfully be quantitatively compared with analogous measurements in real-world data. The ultimate point of these models is often to explain precise qualitative and quantitative results in real-world data by comparison with analogous results in the models.

It is worth emphasizing that a model can explain how some phenomenon occurs only if it produces actual examples of the phenomena in question; it is not sufficient to produce something that represents the phenomenon but lacks its essential properties. In my view, this is the problem with S. J. Gould's "drunkard walk" model of the growth of complexity in the biosphere [6]. In effect, Gould's model is a random branching process producing unstructured entities distributed along a dimension which you can think of as the positive integers. Gould then interprets location on that dimension as morphological complexity, with the corresponding integer's size or order measuring degree of complexity. But that interpretation is totally arbitrary. Nothing produced by the model is any more or less complex than anything else, as you can see by noting that you could just as well interpret the integers as measuring the reciprocal of degree of complexity. So Gould's model sheds no light on how a process produces entities with an interesting distribution of degrees of complexity.

3 Using Unrealistic Models

Hraber et al. [9] distinguish three different ways one might construe the intended use of Echo, Holland's celebrated artificial-life model. It could be construed as (i) something which corresponds directly to some real ecosystem, or as (ii) an ecological abstraction used to build intuitions about the general properties of ecosystems, or as (iii) a general theory of complex adaptive systems. I doubt that any of these exactly matches the use I have in mind for unrealistic models. Unrealistic models are much too abstract to correspond directly to some specific real ecosystem, and I'm skeptical that any single unrealistic model will capture enough of the essential features of enough different systems to serve as a general theory of complex adaptive systems. I do view unrealistic models as abstractions for building intuitions, but Hraber et al. may intend to build precise intuitions about detailed systems; they make an analogy with flight simulators, and these are used to build quite realistic intuitions about flying specific airplanes.

I view unrealistic models as thought experiments, intended to capture the essential properties of some kind of complex adaptive system. One can view Packard's Bugs model [3] and Ray's Tierra model [15] as embodiments of two related but different pictures of the essential core of living evolving systems. The goal of capturing the essential properties of a complex adaptive systems is typically to discern the essential mechanisms underlying certain interesting kinds of phenomena, such as multi-level emergence, open-ended evolution, or complexity growth. Although one may become interested in understanding the mechanisms behind such phenomena out of the conviction that actual biological systems exhibit them, this need not be so. Even though it is controversial whether the biosphere really exhibits open-ended evolution or unbounded complexity growth, for example, one can still seek the simplest way to generate those phenomena. Having such a model help one determine whether those phenomena *are* present in the biosphere.

With many models, and especially with unrealistic models, the model's key emergent properties can be observed only by means of certain macro-level statistics. (Thermodynamic analogues are statistics like pressure and temperature.) Some examples of such statistics are diversity of genotypes [2] or species [13], or "complexity" of taxa (see Ref. [14] and references cited therein), or adaptive evolutionary activity [3, 5]. A statistic enables you to see a model's macro-scale forest in the face of all the micro-scale trees. In addition, these statistics are what allow you to compare the behavior of the model with the behavior of the natural systems you seek to understand. For this reason, it is typically in terms of such statistics that fundamental theories, laws, and classifications would be expressed. Whereas Holland might propose a universally applicable *model* like Echo as the route to a general theory of complex adaptive systems [8], I see universally applicable *statistics* as providing this route. From this point of view, a crucial part of the intellectual effort of using unrealistic computer models is devising statistics to operationalize the key macro-scale concepts, such as emergence, adaptation and complexity.

4 Validating Unrealistic Models

Axtell and Epstein propose that agent-based models should be validated ultimately through quantitative agreement with empirical macro- and micro-structures [1]. But since an unrealistic model is not a model of any particular real systems, this validation methodology might not be relevant. The same holds for Maley's method of validating artificial life models against previous verified theories [12]. Since unrealistic models are thought experiments, it is an open question whether previous theory applies to them. In addition, we may well have no significant theories of the behavior of systems exhibiting deep unexplained phenomena like multiple-level emergence, open-ended evolution, and unbounded complexity growth. Still, unrealistic models do not escape the validation issue since one must ensure it is "valid" to use the models to understand the systems or phenomena of interest.

How could one check whether a given unrealistic model is suitable for illuminating some behavior of some kind of complex biological system? The first thing one needs is some objective and operational method for detecting the behavior in question. Here, the appropriate macro-level statistics can be of great assistance. The next step is to simply compare the behavior of the model and the biological system. When I attempted to make just this kind of comparison between the long-term trends involving adaptive evolution in artificial life models and the biosphere are reflected in the fossil record, I concluded that the models failed to capture the sort of unbounded adaptive creativity apparently evident in the biosphere [4, 5], i.e., the models were *invalid* for illuminating the biosphere's evident unbounded adaptive creativity.

This validation method is more ambiguous than I am suggesting, though. One could criticize my conclusion on the grounds that I let the artificial life models collide with their resource ceilings, while this has not yet happened with the biosphere. Maley attempts to avoid this criticism by focusing only on the initial transient in the model's behavior, before it reaches its carrying capacity [13] (see also the discussion in [4]). But Maley and I share the suspicion that the artificial life models are qualitatively different from the biosphere in that the models lack the capacity for new kinds of niches to be continually created as an intrinsic consequence of the course of adaptive evolution. I know of no recipe for determining the proper conditions under which to test for this sort of property.

Are we entitled to conclude today that it *is* valid to use unrealistic models to answer deep questions about multi-level emergence, open-ended evolution, or complexity growth? To my mind, we do not yet have adequate models of any of these phenomena, though this is not the place to argue the point. Will some unrealistic model illuminate these phenomena in the future? This is an empirical question and we lack the relevant empirical data to discern the answer. In principle the question could have an affirmative answer, but that does not imply that the answer probably *is* affirmative. By the same token, our present lack of evidence for an affirmative answer is not evidence for the lack of an affirmative answer. Dismissing the attempt to answer these questions with unrealistic models would be to fail to recognize that these profound questions are exactly the sort of mysteries that unrealistic models might well explain.

Acknowledgements. For helpful discussion and comments on the manuscript, thanks to Carlo Maley. Thanks also to the Santa Fe Institute, for support and hospitality while some of this work was completed.

References

[1] Axtell, R. L, and J. M. Epstein. 1994. Agent-based modeling: understanding our creations. *The Bulletin of the Santa Fe Institute* 9: 28–32.

[2] Bedau, M. A. 1995. Three illustrations of artificial life's working hypothesis. In W. Banzhaf and F. Eeckman, eds., *Evolution and Biocomputation—Computational Models of Evolution* (pp. 53–68). Berlin: Springer.

[3] Bedau, M. A. and N. H. Packard. 1992. Measurement of evolutionary activity, teleology, and life. In C. Langton, C. Taylor, J. D. Farmer, S. Rasmussen, eds., *Artificial Life II* (pp. 431–461). Redwood City, CA: Addison-Wesley.

[4] Bedau, M. A., E. Snyder, C. T. Brown, N. Packard. 1997. A comparison of evolutionary activity in artificial evolving systems and in the biosphere. In P. Husbands and I. Harvey, eds., *Proceedings of the Fourth European Conference on Artificial Life* (pp. 125–134). Cambridge: MIT Press/Bradford Books.

[5] Bedau, M. A., E. Snyder, N. Packard, 1998. A classification of long-term evolutionary dynamics. In C. Adami, R. Belew, H. Kitano, and C. Taylor, eds., *Artificial Life VI* (pp. 228–237). Cambridge: MIT Press/Bradford Books.

[6] Gould, S. J. 1996. *Full House*. New York: Harmony Books.

[7] Gould, S. J., Lewontin, R. C. 1979. The spandrals of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society B* 205: 581–598.

[8] Holland, J. H. 1992. *Adaptation in Natural and Artificial Systems*, 2nd edition. Cambridge: MIT Press/Bradford Books.

[9] Hraber, P. T., T. Jones, and S. Forrest. 1997. The ecology of Echo. *Artificial Life* 3: 165–190.

[10] Levin, S. A., B. Grenfell, A. Hastings, and A. S. Perelson. 1997. Mathematical and computational challenges in population biology and ecosystems science. *Science* 275: 334–343.

[11] Lindgren, K. 1992. Evolutionary phenomena in simple dynamics. In C. Langton, C. Taylor, J. D. Farmer, S. Rasmussen, eds., *Artificial Life II* (pp. 295–312). Redwood City, CA: Addison-Wesley.

[12] Maley, C. C. 1998. Models in evolutionary ecology and the validation problem. In C. Adami, R. Belew, H. Kitano, and C. Taylor, eds., *Artificial Life VI* (pp. 423–427). Cambridge: MIT Press/Bradford Books.

[13] Maley, C. C. 1999. Four steps toward open-ended evolution. In Banzhaf, W., J. Daida, A. E. Eiben, M. H. Garzon, V. Honavar, M. Jakiela, and R. E. Smith, eds., *GECCO-99: Proceedings of the Genetic and Evolutionary Computation Conference*. San Francisco, CA: Morgan Kaufmann.

[14] McShea, D. W. 1966. Metazoan complexity: is there a trend? *Evolution* 50: 477–492.

[15] Ray, T. S. 1992. An approach to the synthesis of life. In C. Langton, C. Taylor, J. D. Farmer, S. Rasmussen, eds., *Artificial Life II* (pp. 371–408). Redwood City, CA: Addison-Wesley.

[16] Rechtsteiner, A. and M. A. Bedau. 1999. A generic neutral model for measuring excess evolutionary activity of genotypes. In Banzhaf, W., J. Daida, A. E. Eiben, M. H. Garzon, V. Honavar, M. Jakiela, and R. E. Smith, eds., *GECCO-99: Proceedings of the Genetic and Evolutionary Computation Conference*. San Francisco, CA: Morgan Kaufmann.