

Book Review: Search and Construct

Susan Stepney*
University of York

Artificial Chemistries. Wolfgang Banzhaf and Lidia Yamamoto. (2015, MIT Press.) 576 pages.

An enormous quantity may be termed “astronomical,” referencing the huge span of time since the Big Bang ($\sim 10^{17}$ s), the huge size of the universe ($\sim 10^{27}$ m), or the huge amount of material in the observable universe ($\sim 10^{80}$ atoms). Yet these quantities pale into insignificance compared to those generated by combinatorics, where numbers are combined using multiplication and exponentiation, leading to an explosion in their size. The number of possible proteins of the typical length of eukaryotic proteins is $20^{400} \sim 10^{520}$ (although not all of these would have a sensible shape or function); the number of possible memory configurations of a mere 1 kB of RAM is $2^{8 \times 2^{10}} \sim 10^{10^3}$; the number of books in Borges’ *Library of Babel* is more than 10^{10^6} (yet hardly any are interesting books), they can be shelved in $\sim (10^{10^6})$ to the power 10^{10^6} ways, and even the library’s catalogue is huge; and so on.

Daniel Dennett, in his book *Darwin’s Dangerous Idea*, uses a clever trick to remind us of the sheer scales involved. He builds up an intuition, or possibly more of a feeling, of such sizes, then dubs these “Vast,” with a capital V. Ever after, the term Vast evokes that sheer scale.

Within the Vastness of all possibilities, only a subset is somehow interesting: Most is mere noise. This subset may be Vast in its own right, yet vanishingly small relative to the Vastness of all possibilities. How to find such vanishingly small needles in the Vastness of a combinatoric haystack?

One technique might be dubbed “search and construct.” Search for a useful set of atoms, primitives, components, that form the basis of the Vast combinatorial space. Then use rules and processes to define or generate only those constructs with interesting structure and behavior within that space. For the Library of Babel, the atoms are characters, and the Vastness is all possible books of these characters. But what rules delimit the subspace of *interesting* books, books that are grammatical, readable, and worthwhile? There is chunking to form higher-level components: words. There are syntactic restrictions on the form of sentences, and further semantic restrictions to be meaningful. But to go further, to construct the subset that is *literature*, say, requires as yet uncodified human creativity. For computer programming, constructing a member of the interesting subset is a slightly easier task. The primitives are the relevant high-level language constructs and their syntactic constraints, the rules include well-formedness constraints and patterns, yet there is still much creativity needed to construct useful programs.

Many researchers turn to the natural world for inspiration. Evolution is one process that explores these interesting possibilities. It can be considered part of a process that searches for genomes, then constructs phenotypes. Interestingness here is viability. A range of artificial evolutionary algorithms take inspiration from these natural processes. In nature, the starting point for evolution is already something quite complex: An organism, even a single-celled organism, is nontrivial, not a random collection of molecules. Can we find a mechanism for generating this initial complexity?

Underlying life is chemistry. Chemistry is combinatorics *par excellence*. From a small set of atoms, chemical bonding laws produce a Vast set of molecules with structure and behavior. It has chunking: Atoms can form small molecular building blocks, such as DNA bases and amino acids, that are themselves the components in higher-level constructions. Good blocks can be searched and selected for by evolution. As we have seen from the protein example above, the larger molecules produced are still a vanishingly small subset of the potential Vastness. Not all combinations of atoms can form

* Department of Computer Science, Heslington, University of York, York YO10 5DD, UK.

stable molecules, and not all molecules that can form have a function or structure that can contribute to further construction.

Artificial chemistry (AChem) takes such ideas from natural chemistry, in order to generate and explore a variety of forms of combinatoric Vastness *in silico*. If we think of AChems as a generic form of “search and construct” processes, and as rule-based novelty generators, we can see that they can be applied not simply to chemical problems, but to a whole range of domains where such processes are needed and used, including computing, dynamical systems, language and music, and modeling *in silico* and *in vitro* complex systems.

An AChem provides three components for virtual world explorations. First, there is the *material*, the virtual atoms and molecules, that provides the Vast combinatorial space of potential structures. Then there are the *reaction rules*, the analogs of the laws of nature in our virtual world, which define how the material combines and dissociates, and possibly even how the space it occupies is restructured (as with P-systems). These rules implicitly define a subspace of possible structures in that Vastness. Finally, there is the *algorithm*, which lays out our explicit experimental setup to explore that implicit subspace, anywhere from exhaustive search to pouring some virtual stuff in a virtual bucket and watching what happens.

Nature provides just the one particular kind of material (real-world atoms and molecules) and one set of rules (chemical bonding and reactions) that say which molecules are possible, and which are not. The only freedom the scientist has is in the algorithm: the experimental setup that controls which molecules encounter which others, under what environmental conditions. Despite its real-world constraints, chemistry provides all the richness and complexity sufficient for life itself.

The playpen of AChem is even richer, since we also have the freedom to choose different basic material, and different rules. Yet it has the corresponding downside in that we now have to *implement* the rules of our virtual world.

This new book forms a comprehensive introduction to many different facets of the discipline of AChem. The plurality in its title, *Artificial Chemistries*, indicates the diversity of approaches covered. It covers the why, how, and what of the choices of material, rule, and algorithm, and their consequences. For the beginning student, it provides a wide-ranging review of the subject, and its 1000-item bibliography is a marvelous resource in its own right, providing entry into the relevant scientific literature. For the practicing AChemist, it provides an invaluable reference material on all topics in the discipline.

Despite its comprehensive nature, this book is no mere annotated bibliography: Its structure provides a narrative unity for the discipline. Part I comprises four foundational chapters, laying out the philosophy and scope of the subject, illustrated with some simple example AChems. It includes a primer on basic concepts from chemistry, such as chemical reactions, the law of mass action, equilibrium, chemical bonds, and catalysis. It also covers differential equation modeling and computational techniques.

Part II comprises four chapters covering the natural-world inspiration. It starts with the chemistry of life: that of biochemistry and large organic molecules, including proteins, RNA, and DNA. The level of detail is useful for showing the underlying complexity and richness of the chemical processes that are frequently abstracted as mere string concatenation. It would probably do students good to review this material again once they have designed their initial AChem, to help them appreciate the simplifications they have made. The next chapter discusses simple cells, including their structure with lipid walls, and their dynamics in terms of metabolism. It includes discussion of autopoiesis, Robert Rosen’s ideas on organization in living systems, origin-of-life theories, and more. All this is necessarily brief, as each topic has deservedly book-length treatment elsewhere, and so things can get quite dense in places: The Rosen section in particular will probably be incomprehensible to anyone who has not already encountered the material. But the bibliography will guide the curious reader to further explanations. Next come chapters on evolution and open-ended systems. Open-endedness is the holy grail of AChems: Not only can they explore a Vast configuration space, they may be able to grow this very space by opening up new possibilities and dimensions through their own contingent development. These chapters contain a mix of fairly standard material given added

value by being filtered through an AChem perspective—for example, evolutionary dynamics is discussed in terms of chemical reactions—and some quite deep and provocative concepts.

Part III comprises three chapters of massive literature review, documenting AChems and categorizing them into rewriting systems, automata, and bio-inspired. In rewriting systems the reaction rules state how a particular string or other representation is systematically changed into a new form; these include lambda calculi, P-systems, L-systems, and the like. Automata AChems comprise molecules whose atoms are assembly-language-level computational instructions: Molecular behavior is given by the execution of these fragments. These include specific systems such as Tierra and Avida, as well as more generic systems such as cellular automata, von Neumann constructors, and all the way up to Turing machines. The bio-inspired AChems hold more closely to biological mechanisms, such as enzyme reactions, RNA binding, shape-based lock-and-key binding, genetic networks, and swarms. These chapters demonstrate a strength and weakness of AChems: the ability to build yet another arbitrary complex system. Some of these AChems have been examined in detail over a long period of time by research groups; others exist in only an article or two resulting from a single doctoral student project. These chapters can be used as a reference to find specific AChems, or as source material for developing new AChems, hopefully as a synthesis and unification of existing ones. Their comprehensive nature can be a problem on occasion: A whole algorithm may be covered in a single spare sentence. Yet the vast bibliography leads on to more detail.

Part IV comprises four chapters focusing on the global dynamics of general AChems. While parts II and III will be best for students, this part will be of most value to more experienced researchers. First is a chapter on organization theory, written with Pietro Speroni di Fenizio. This looks at conditions for and properties of closed sets of molecules: sets where each molecule is produced by members of the set, and so the reaction network is closed. The following chapter discusses the dynamics of such organizations: effects of reaction rates and probabilities on their construction and maintenance. Next comes a chapter dealing with what for me is the *raison d'être* of AChems: emergence. It provides a discussion of relevant topics: self-organization, nonequilibrium thermodynamics, chaos, downward causation, all as they are relevant to AChems. Several deep and important concepts are each outlined in half a page, and the chapter covers a stunning range of topics. The final chapter in this part continues the theme of emergence by discussing constructive dynamical systems: how AChems can produce novelty.

Part V comprises five chapters on applications of AChems to a wide range of domains. Here we get discussion of everything from robotics to unconventional computation, from nuclear physics to economics, from modeling biological systems to synthetic biology.

The book also includes an appendix giving details of the PyCell AChem package, which provides an immediate entry to computational AChems.

This book is really three or more significant books rolled into one, as needed to cover the breadth of the subject. There are interdisciplinary issues here: A practitioner needs to know a lot about a wide range of subjects. As such, it is a remarkable work of scholarship, bringing together a whole host of diverse information, and synthesizing it into a coherent and valuable account of the discipline of artificial chemistry. I learned a lot from reading it—not just the material that was new to me, but also new ways of looking at known material, and the valuable syntheses of a wide range of concepts. The authors should be commended for their impressive contribution to the field. Any AChemist, ALifer, or, more generally, nature-inspired computer scientist or engineer will find *Artificial Chemistries* a valuable addition to their research bookshelf.