THE INVENTION OF AN ALGORITHMIC BIOLOGY

Seth Bullock

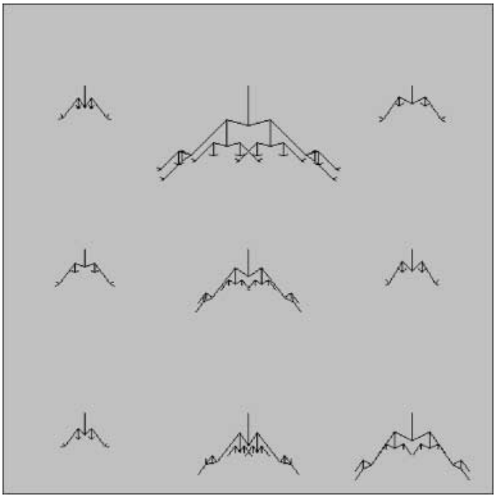
Biology and computing might not seem the most comfortable of bedfellows. It is easy to imagine nature and technology clashing as the green-welly brigade rub up awkwardly against the back-room boffins. But collaboration between the two fields has exploded in recent years, driven primarily by massive investment in the emerging field of bioinformatics charged with mapping the human genome. New algorithms and computational infrastructures have enabled research groups to collaborate effectively on a worldwide scale in building huge, exponentially growing genomic databases, to 'mine' these mountains of data for useful information, and to construct and manipulate innovative computational models of the genes and proteins that have been identified. This recent burst of high-profile activity might suggest that computer scientists have only recently begun to work on biological questions, but activity at this particular disciplinary interface is by no means new. In fact, it has an extremely long history involving the most famous early pioneers of computing, cybernetics, and artificial intelligence.

In the 1950s, Alan Turing, the 'father of artificial intelligence' and a man fundamentally associated with codes, logic, chess, and other mechanico-mathematical arcana, developed influential models of biological morphogenesis:[1] the processes involved in the development of biological patterns as an organism grows from a single cell. He was particularly interested in accounting for the tendency of spiral patterns in many plant structures to obey the Fibonacci sequence (e.g. if you count the number of whirls running clockwise on a pine cone and the number running anticlockwise, the two numbers will be consecutive terms in Fibonacci's famous sequence of integers: 0, 1, 1, 2, 3, 5, 8, 12, . . .). At the same time, John von Neumann, one of history's great polymaths and the man responsible for game theory and the architecture of the modern computer among many other things typically considered to lie far from the muddy field of biology, worked on the problem of self-replication:[2] over evolutionary time, simple life-forms have given rise to more complicated creatures, but how, von Neumann asked, could a machine (like a dog or an amoeba or a robot) make a more complex version of itself? The answer that he arrived at predicted the essential distinction between DNA (instructions) and transcriptase (machinery that follows instructions) several years before Crick and Watson's discovery.

Surprisingly, though, the very first example of activity fusing computing and biology is over a century older than the work of Turing and von Neumann, predating even Darwin's *Origin of Species*. It is due to Charles Babbage, designer of the Difference Engine, the first automatic calculating machine and the progenitor of the modern computer. As early as 1837, Babbage reported using this machine to help him demonstrate that inexplicably abrupt changes in the geological record need not be taken to be the work of God (a hot topic of the day). He showed that his completely deterministic (clockwork) machine could generate analogous surprising behaviour ('miracles') without any external interference from the programmer. He invited his contemporaries (including Darwin) to observe the machine generating a sequence of numbers (1, 2, 3, 4, . . .) and asked them to state the rule or law that the machine was obeying. At some predetermined point the engine would 'disobey' this law, automatically beginning to generate some alternative stream of numbers (the Fibonacci sequence, perhaps), and surprising the onlookers, who were forced to admit that apparently mysterious and abrupt changes observed in nature need not demand explanation in terms of divine intervention.[3]

It is clear, then, that computing and biology have communicated from almost the first possible moment, and have been finding new and productive ways to interact ever since. And it is firmly within this tradition that some of Richard Dawkins' most interesting work can be located. Indeed, there are two senses in which this is true. First, and most straightforwardly, Dawkins has had a significant involvement in the development of *bio-inspired algorithms*, specifically within the field of evolutionary computation, where computer programs solve problems in a manner inspired by biological evolution. In 1986, in *The Blind Watchmaker*, Dawkins introduced an algorithm of the same name. This computer program requires a user repeatedly to select one of nine bilaterally symmetrical line drawings, or 'biomorphs' (see Fig. 1). After each selection, nine new variants of the chosen biomorph are randomly generated and presented. Over time, the line drawings 'evolve' to reflect the taste of the user, who is effectively breeding biomorphs by exerting selection pressure on a population of forms that are competing with one another for the chance to 'reproduce'.

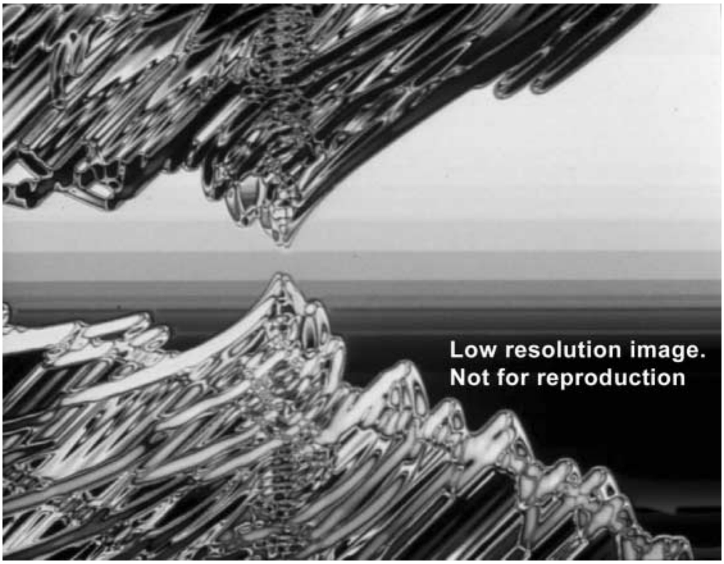
A year later, Dawkins presented his biomorphs at an 'Interdisciplinary Workshop on the Synthesis and Simulation of Living



1. Biomorphs 'evolved' using the *Blind Watchmaker* program.

Systems' held at Los Alamos National Laboratory in New Mexico. The meeting brought together a disparate group of researchers from computing, mathematics, physics, biology, neuroscience, and even economics to talk about a set of topics that have come to be known collectively as Artificial Life.[4] What is life? Can it be synthesized *in silico*? What can we learn about life in the attempt? Dawkins' involvement at the outset of artificial life (along with that of other biologists such as Elliot Sober and John Maynard Smith) leant the field some credibility, but his contributions[5] to the first conference is also notable in its own right. In it, he presented the *Blind Watchmaker* program as a tool with which to explore the notion of *evolvability*—the tendency of a population to tolerate and eventually profit from small changes (mutations). This property remains poorly understood. While biological progeny are not identical to their parents or their siblings, they typically remain viable organisms. By contrast, introducing a few random mutations into a computer program or a hospital's working procedures is likely to prove catastrophic. Moreover, the mutations suffered by biological organisms are not just neutralized, corrected, or ironed out, since enough useful variation amongst relatives remains to fuel natural selection. This balance between robustness and sensitivity, between staying the same and changing, has yet to be understood and exploited in evolutionary computation or other relevant fields—amongst other things, a full understanding of it would revolutionize our ability to manage evolving complex systems such as hospitals, cities, economies, and so on. Dawkins' paper represents an early attempt to address some of these issues.

Dawkins' program itself is unusual in that, unlike standard evolutionary algorithms, it demands that the user *manually* exert selection pressure on an artificial evolving population, choosing which 'biomorphs' get to reproduce. This approach has inspired a whole *oeuvre* of 'aesthetic evolutionary algorithms' in which artists produce their art in *partnership* with an artificial evolutionary process, moving far beyond Dawkins' stick figures, to generate much more complex pieces[6] (see Figs. 2 and 3). Our commonsense



2 and 3. Evolved artwork. © Karl Sims, used by kind permission.

notion of artistic creativity combines both a *generative* aspect (actually making, altering, improving the artefact) with a *selective* aspect (choosing whether the alteration makes the artefact better, or complete). By contrast, Dawkins' evolutionary approach cedes responsibility for generation to the computer which randomly (rather than purposively) perturbs the currently selected individual. The artist reserves only the right to sift these perturbed forms and select which of them are to be (mis)copied into the next generation.

As such, in addition to serving as a tool with which to introduce adaptation by natural selection to a general audience, the program raises a number of interesting questions concerning progress, purpose, and creativity in art and nature. Is the user of such a computer program really an artist, and if so what is the status of the program's writer? Can pointing and clicking one's way through a (potentially infinite) genetic space of 'predefined' forms be somehow equivalent to painting or drawing? In fact, the practice resembles the (non-artistic?) selective breeding of plants, livestock, or domestic animals, but simultaneously resonates with some experimental art in which the artist's volition is similarly attenuated (e.g. Jackson Pollock's action painting, which coupled spontaneous 'random' splashing and dripping with careful subsequent editing, cropping, or outright rejection).[7]

While Dawkins' simple computer program was the first example of a commercially released piece of artificial life software, its potency is better evidenced by the number of times it has been recoded and extended by those who have read about it. The internet is home to a veritable cottage industry of biomorph breeding, and many programmers (including my teenage self) must have written their variants of the *Blind Watchmaker* before the internet allowed them to be widely disseminated. There is something compelling in the combination of simplicity, scope, and visual impact that captures the imagination of these programmers, and comes to influence the way that they think about evolutionary processes and algorithms. This is the second sense in which Dawkins' work lies at the boundary between computing and biology—the pedagogical use of specific algorithms and algorithmic thinking and talking to understand and explain evolutionary biological processes: what might be termed algorithmic biology.[8]

An algorithm is a set of step-by-step instructions, like a cake recipe or travel directions. As such, our tacit understanding is that they are useful, but inert and straightforward. Dawkins employs an algorithmic device explicitly when he describes, in *The Blind Watchmaker*, how a particular string of symbols (the sentence: 'Methinks it is like a weasel') might arise via reproduction, mutation, and selection in a population of initially random symbol strings.[9] By repeatedly applying the same sequence of actions, the appearance of deliberate design is achieved despite the randomness inherent to the process and the vast number of possible sentences (roughly 27^{28} if we don't care about upper case or punctuation). Like his biomorphs program, this algorithm is a powerful rhetorical device because it mechanizes and thereby demystifies natural selection (at the expense, perhaps, of muddying the waters concerning the nature of biological selection pressures, which are neither aesthetic nor aiming at a prearranged target).

There are, of course, alternative ways of conveying the central tenets of natural selection: drawing parallels with selective breeding of pigeons or flowers; conveying the impact of finite resources on heritable variation; explaining the implications of the second law of thermodynamics for copying processes. Dawkins makes use of many of these, but algorithmic devices are special. One of their key features is that they are *multiply realizable*. This just means that the same algorithm can be carried out by many different machines. You or I could follow the same set of directions and your computer or mine could execute the *Blind Watchmaker* program (if the languages in which the algorithms are written are appropriate). Algorithms abstract away from nitty-gritty implementation details (just where is my cake tin? how exactly do I 'jump on the No. 1 bus'?), casting a process at a level that rises somewhat above particular instances of execution, without resorting to mathematical or logical formalisms that have limited currency.

Crucially, when taking an algorithmic approach to natural selection, rather than writing in terms of, for example, competition for scarce resources (fighting, fleeing, feeding, sex), the evolutionary process is free to dissociate from the 'four Fs', thereby becoming readily applicable to a wider range of non-genetic (quasi-)evolutionary systems. Most famously, and much earlier, in *The Selfish Gene* Dawkins was able to reapply the abstracted principles of natural selection within the realm of ideas, conjuring the *meme* as an ideational equivalent of the biological gene.[10] Since then, there has been a significant proliferation of (quasi-) evolutionary approaches to a range of non-genetic systems: evolutionary linguistics, economics, psychology, and even cosmology, as well as evolutionary computation and art. In most cases, the success or failure of these enterprises cannot yet be judged, but their very existence is testament to the expanding power of uprooted evolutionary biological concepts and, in particular, the biological algorithm at the heart of evolution by natural selection.

The pioneers name-checked at the outset of this paper suggest that, historically, most significant workers at the computing–biology interface have tended to be mathematicians or computer scientists who are interested in biological questions. Dawkins bucks this trend somewhat, in that he is a biologist, and one who has not been particularly interested in computational questions. Rather, he is interested in *using* computers, not just as tools with which to write or calculate, but primarily as tools with which to think.

ENDNOTES

1 A. M. Turing, 'The chemical basis of morphogenesis', *Philosophical Transactions of the Royal Society of London, Series B*, 237 (1952): 37–72.

2 J. von Neumann and A. W. Burks, *Theory of Self-Reproducing Automata* (Urbana, IL: University of Illinois Press, 1966).

3 Babbage's automated 'miracles' did, however, suggest an image of God as an omnipotent engineer, having no need for divine intervention, but instead relying upon 'pre-programmed' miraculous events. At the time, it remained an act of faith to believe that this kind of pre-programming could, unlike fundamentally inexplicable acts of divine intervention, be understood through regular science (in the same way that careful examination of Babbage's engine would have revealed the laws that governed its surprising behaviour). For more discussion, see S. Bullock, 'Charles Babbage and the emergence of the evolutionary simulation model', in P. Husbands, O. Holland, and M. Wheeler (eds.), *The History of the Mechanization of Mind* (Cambridge, MA: MIT Press, forthcoming).

4 C. G. Langton (ed.), *Artificial Life: An Overview* (Cambridge, MA: MIT Press, 1995).

5 R. Dawkins, 'The evolution of evolvability', in C. G. Langton (ed.), *Artificial Life: The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems* (Redwood City, CA: Addison-Wesley, 1989).

6 K. Sims, 'Artificial evolution for computer graphics', *Computer Graphics*, 25/4 (1991): 319–332.

7 A. Dorin, 'Aesthetic fitness and artificial evolution for the selection of imagery from the mythical infinite library', in J. Kelemen and P. Sosík (eds.), *Proceedings of the Sixth European Conference on Artificial Life* (Berlin: Springer, 2001), 659–668.

8 I owe the term to Richard A. Watson, although my usage may not be equivalent to his.

9 R. Dawkins, *The Blind Watchmaker* (New York: W. W. Norton, 1986).

10 R. Dawkins, *The Selfish Gene* (Oxford: Oxford University Press, 1976).