Proving Darwin: Making Biology Mathematical

by Gregory Chaitin

reviewed by Jeffrey Shallit

This is an infuriating little book.

It’s poorly written—if the author knows what a run-on sentence is, he doesn’t care. It’s repetitive and padded with a thirteen-page paper of John von Neumann (1968) that is easily found on the web (for example, at http://tinyurl.com/7kwxgud), so that the book’s 123 pages suggest more content than is really there. The author, Gregory Chaitin, is relentlessly self-promoting, and likes to drop the names of his famous and infamous friends, including Stanislaw Ulam, Jack Schwartz, Sydney Brenner, Stephen Wolfram, Marvin Minsky, and David Berlinski (a senior fellow of the Discovery Institute’s Center for Science and Culture!). He portrays himself as a scientific rebel, and favorably cites the bogus cold fusion claims of Italian inventor Andrea Rossi (Wolchover 2012). The book makes some false claims about biologists and exaggerated claims about the importance of the mathematical results it contains, and Chaitin takes sole credit for a theory that was largely developed earlier by others.

Nevertheless, despite all these flaws—and more problems, which there is not enough space to describe here—the book is written in an engaging and enthusiastic style, and does contain one rather interesting idea, which I will explain below. Readers interested in pursuing more details can easily find Chaitin’s papers on the web (Chaitin 2010, 2011, 2012), so there is really no good reason to buy this book.

First, let me tell you a little bit about the author. In 1965, at the age of 18, while still an undergraduate at City College in New York, Chaitin submitted two papers on the foundations of what is now called “Kolmogorov complexity” or “algorithmic information theory”; one was published in 1966 and the other in 1969. Chaitin has spent the rest of his career mostly on working out the implications of these ideas.

However—and this is common in the history of mathematics and science—the basic ideas of algorithmic information theory had been discovered earlier by Ray Solomonoff and Andrei Kolmogorov. Because of priority, and because Kolmogorov was already world-famous for his work in probability theory, the name “Kolmogorov complexity” became entrenched in the literature for this field. Chaitin certainly deserves credit for his independent discovery, especially at so early an age, and since then he has found many additional results of interest, such as his amazing number, Omega (Gardner 1979). However, while algorithmic information theory is discussed briefly in the book under review, the reader will not find the theory’s co-inventors Solomonoff and Kolmogorov mentioned anywhere. This is unfortunate. (For an unbiased appraisal of each person’s relative contribution, see Li and Vitanyi 1997:89–92.)
But back to the main subject of the book. For many years people have attempted to model biological evolution using computer programs. Indeed, there is an entire field devoted studying this, called “artificial life”, with yearly conferences and an academic journal published by MIT Press. (Oddly enough, the term doesn't seem to appear anywhere in the book under review.) However, a significant emphasis in artificial life thus far has been the construction of software models of organisms that evolve in various ways, such as Tom Ray's Tierra (Ray nd), and Karl Sims's evolution of locomotion strategies (1994a, 1994b). In this milieu, experiments are key, and proof of a claim consists of coding up your simulation, running it, and seeing what you get.

Chaitin's approach is very different. He wants to construct a model of evolution that permits one to rigorously prove mathematical theorems about what happens in the model.

Chaitin's model of evolution is so simple it can be described in a couple of paragraphs. An organism is modeled by a computer program $P$ that, when you run it, must eventually halt and print out an integer $I$. At each step of the simulation, a mutation is applied to $P$, obtaining a new program $P'$ that prints out some other integer $I'$. If $I'$ is bigger than $I$, the new program $P'$ is deemed to be more fit; $P$ is then killed off and $P'$ replaces it. For Chaitin, this is how evolution proceeds.

How are Chaitin's mutations applied? Randomly and algorithmically. More precisely, the mutations themselves are also computer programs, and at each step, every possible mutation program is considered, with a probability distribution that strongly favors “simple” mutation programs being chosen and disfavors more complicated ones. Specifically, we assign a certain probability of being chosen to every program of length $n$, for all integers $n$. At each step we choose such a program $Q$ according to our probability distribution, run $P$ through it, and get a new program $P' = Q(P)$. This is the model.

Chaitin's main result, which he can rigorously prove, is that in such a model, after at most $n^2(\log n)^2$ steps, we will obtain, with high probability, a program to compute the busy-beaver function $BB(n)$, which is the function sending $n$ to the largest possible integer printable by a $n$-bit program. The significance of the evolution, in Chaitin's model, of programs that compute the busy-beaver function lies in the fact that busy-beavers are hard to find, both because the search space (of all possible $n$-bit programs) is so large and because it is in general impossible to determine whether an arbitrary program will halt. He contrasts this result with a model based on pure random choice, which will take about $2^n$ steps to get the same result; this is much longer than his evolutionary model. Finally, in a model he calls “intelligent design”—which means that at each step, the optimal mutation is chosen—about $n$ steps suffice.

A pure mathematician or theoretical computer scientist may well find this result interesting—and I do! But if it is supposed to be relevant to biological evolution, there are a number of obvious objections. First, when we get $P'$ by mutating $P$, it could well be that $P'$ is not a useful computer program. It could, for example, go into an infinite loop and never print out anything at all. And of course, we know from fundamental results of Alan Turing that there is no program that will, in general, tell us whether $P'$ is useful in this sense. To get around this, Chaitin blithely assumes we have an “oracle”: a purely theoretical construct in computer science that ignores Turing’s theorem and gives answers to uncomputable
problems. We have a similar problem when we run our mutation program Q, so we need an oracle to handle that, too.

But as soon as we assume we have an oracle to solve the halting problem, we could solve the busy beaver problem directly, simply by running all programs of length N, and weed¬ing out those that don't halt! To get around this, Chaitin needs yet another rule: the oracles can't be used arbitrarily, but only in the way he specifies. The result is that his model is rather arbitrary and says little, if anything about the actual physical process of evolution.

Furthermore, Chaitin's model is unrealistic in other ways. Evolution occurs in populations, not single individuals. In a population we get competition for scarce resources, and we can get sex and horizontal transfer. Furthermore, mutations do not seem to be algorithmic in Chaitin's sense; they seem to be restricted to a few very basic kinds of changes, such as point mutations.

Then again, Chaitin is not a biologist. Readers will be astonished to learn that “every cell in our body has the complete DNA for an entire human being” (p 17); I guess those red blood cells and gametes aren't worth chopped liver. As well, we are told that “conventional biologists ... suspect that life on Earth was either seeded by accident ... or deliberately planted” (p 14). Funny, I know a lot of conventional biologists who don't suspect either one.

So contrary to the title of his book, Chaitin has not proved Darwin mathematically. Nor has he invented a whole new field, “metabiology”; indeed, papers by Nehaniv and Rhodes (1997, 1999, 2000) have already explored some analogous ideas. Instead, he's created an unrealistic but intriguing mathematical model quite divorced from biological evolution in the real world, and proved some theorems about it. This is the stuff of an interesting talk at a departmental colloquium, not a whole book.

References


ABOUT THE AUTHOR

Jeffrey Shallit is a professor of computer science at the University of Waterloo. He has written three books and about 130 papers on mathematics and computer science. He was asked to testify in the Dover trial as a rebuttal witness against William Dembski. He blogs at http://recursed.blogspot.com.

AUTHOR’S ADDRESS

Jeffrey Shallit
School of Computer Science
University of Waterloo
Waterloo, ON N2L 3G1 Canada
shallit@cs.uwaterloo.ca

Copyright 2013 by Jeffrey Shallit; licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. http://creativecommons.org/licenses/by-nc-nd/3.0/