

ALGORITHMIC INFORMATION & EVOLUTION

This is a revised version of a lecture presented April 1988 in Paris at the International Union of Biological Sciences Workshop on Biological Complexity. Published in O. T. Solbrig and G. Nicolis, *Perspectives on Biological Complexity*, IUBS Press, 1991, pp. 51–60.

G. J. Chaitin

*IBM Research Division, P.O. Box 218
Yorktown Heights, NY 10598, U.S.A.*

Abstract

A theory of information and computation has been developed: “algorithmic information theory.” Two books [11–12] have recently been published on this subject, as well as a number of nontechnical discussions [13–16]. The main thrust of algorithmic information theory is twofold: (1) an information-theoretic mathematical definition of random sequence via algorithmic incompressibility, and (2) strong information-theoretic versions of Gödel’s incompleteness theorem. The

halting probability Ω of a universal Turing machine plays a fundamental role. Ω is an abstract example of evolution: it is of infinite complexity and the limit of a computable sequence of rational numbers.

1. Algorithmic information theory

Algorithmic information theory [11–16] is a branch of computational complexity theory concerned with the size of computer programs rather than with their running time. In other words, it deals with the difficulty of describing or specifying algorithms, rather than with the resources needed to execute them. This theory combines features of probability theory, information theory, statistical mechanics and thermodynamics, and recursive function or computability theory.

It has so far had two principal applications. The first is to provide a new conceptual foundation for probability theory based on the notion of an individual random or unpredictable sequence, instead of the usual measure-theoretic formulation in which the key notion is the distribution of measure among an ensemble of possibilities. The second major application of algorithmic information theory has been the dramatic new light it throws on Gödel’s famous incompleteness theorem and on the limitations of the axiomatic method.

The main concept of algorithmic information theory is that of the program-size complexity or algorithmic information content of an object (usually just called its “complexity”). This is defined to be the size in bits of the shortest computer program that calculates the object, i.e., the size of its minimal algorithmic description. Note that we consider computer programs to be bit strings and we measure their size in bits.

If the object being calculated is itself a finite string of bits, and its minimal description is no smaller than the string itself, then the bit string is said to be algorithmically incompressible, algorithmically irreducible, or algorithmically random. Such strings have the statistical properties that one would expect. For example, 0’s and 1’s must occur with nearly equal relative frequency; otherwise the bit string could be compressed.

An infinite bit string is said to be algorithmically incompressible, algorithmically irreducible, or algorithmically random if all its initial

segments are algorithmically random finite bit strings.

A related concept is the mutual algorithmic information content of two objects. This is the extent to which it is simpler to calculate them together than to calculate them separately, i.e., the extent to which their joint algorithmic information content is less than the sum of their individual algorithmic information contents. Two objects are algorithmically independent if their mutual algorithmic information content is zero, i.e., if calculating them together doesn't help.

These concepts provide a new conceptual foundation for probability theory based on the notion of an individual random string of bits, rather than the usual measure-theoretic approach. They also shed new light on Gödel's incompleteness theorem, for in some circumstances it is possible to argue that the unprovability of certain true assertions follows naturally from the fact that their algorithmic information content is greater than the algorithmic information content of the axioms and rules of inference being employed.

For example, the N -bit string of outcomes of N successive independent tosses of a fair coin almost certainly has algorithmic information content greater than N and is algorithmically incompressible or random. But to prove this in the case of a particular N -bit string turns out to require at least N bits of axioms, even though it is almost always true. In other words, most finite bit strings are random, but individual bits strings cannot be proved to be random [3].

Here is an even more dramatic example of this information-theoretic approach to the incompleteness of formal systems of axioms. I have shown that there is sometimes complete randomness in elementary number theory [11, 13, 15–16]. I have constructed [11] a two-hundred page exponential diophantine equation with the property that the number of solutions jumps from finite to infinite at random as a parameter is varied.

In other words, whether the number of solutions is finite or infinite in each case cannot be distinguished from independent tosses of a fair coin. This is an infinite amount of independent, irreducible mathematical information that cannot be compressed into any finite number of axioms. I.e., essentially the only way to prove these assertions is to assume them as axioms!

This completes our sketch of algorithmic information theory. Now

let's turn to biology.

2. Evolution

The origin of life and its evolution from simpler to more complex forms, the origin of biological complexity and diversity, and more generally the reason for the essential difference in character between biology and physics, are of course extremely fundamental scientific questions.

While Darwinian evolution, Mendelian genetics, and modern molecular biology have immensely enriched our understanding of these questions, it is surprising to me that such fundamental scientific ideas should not be reflected in any substantive way in the world of mathematical ideas. In spite of the persuasiveness of the informal considerations that adorn biological discussions, it has not yet been possible to extract any nuggets of rigorous mathematical reasoning, to distill any fundamental new rigorous mathematical concepts.

In particular, by historical coincidence the extraordinary recent progress in molecular biology has coincided with parallel progress in the emergent field of computational complexity, a branch of theoretical computer science. But in spite of the fact that the word "complexity" springs naturally to mind in both fields, there is at present little contact between these two worlds of ideas!

The ultimate goal, in fact, would be to set up a toy world, to define mathematically what is an organism and how to measure its complexity, and to prove that life will spontaneously arise and increase in complexity with time.

3. Does algorithmic information theory apply to biology?

Can the concepts of algorithmic information theory help us to define mathematically the notion of biological complexity?

One possibility is to ask what is the algorithmic information content of the sequence of bases in a particular strand of DNA. Another possibility is to ask what is the algorithmic information content of the

organism as a whole (it must be in discrete symbolic form, e.g., imbedded in a cellular automata model).

Mutual algorithmic information might also be useful in biology. For example, it could be used for pattern recognition, to determine the physical boundaries of an organism. This approach to a task which is sort of like defining the extent of a cloud, defines an organism to be a region whose parts have high mutual algorithmic information content, i.e., to be a highly correlated, in an information-theoretic sense, region of space.

Another application of the notion of mutual algorithmic information content might be to measure how closely related are two strands of DNA, two cells, or two organisms. The higher the mutual algorithmic information content, the more closely related they are.

These would be one's initial hopes. But, as we shall see in reviewing previous work, it is not that easy!

4. Previous work

I have been concerned with these extremely difficult questions for the past twenty years, and have a series of publications [1–2, 7–13] devoted in whole or in part to searching for ties between the concepts of algorithmic information theory and the notion of biological information and complexity.

In spite of the fact that a satisfactory definition of randomness or lack of structure has been achieved in algorithmic information theory, the first thing that one notices is that it is not ipso facto useful in biology. For applying this notion to physical structures, one sees that a gas is the most random, and a crystal the least random, but neither has any significant biological organization.

My first thought was therefore that the notion of mutual or common information, which measures the degree of correlation between two structures, might be more appropriate in a biological context. I developed these ideas in a 1970 paper [1], and again in a 1979 paper [8] using the more-correct self-delimiting program-size complexity measures.

In the concluding chapter of my Cambridge University Press book [11] I turned to these questions again, with a number of new thoughts,

among them to determine where biological questions fall in what logicians call the “arithmetical hierarchy.”

The concluding remarks of my 1988 *Scientific American* article [13] emphasize what I think is probably the main contribution of the chapter at the end of my book [11]. This is the fact that in a sense there is a kind of evolution of complexity taking place in algorithmic information theory, and indeed in a very natural context.

The remaining publications [2, 7, 9–10, 12] emphasize the importance of the problem, but do not make new suggestions.

5. The halting probability Ω as a model of evolution

What is this natural and previously unappreciated example of the evolution of complexity in algorithmic information theory?

In this theory the halting probability Ω of a universal Turing machine plays a fundamental role. Ω is used to construct the two-hundred page equation mentioned above. If the value of its parameter is K , this equation has finitely or infinitely many solutions depending on whether the K th bit of the base-two expansion of Ω is a 0 or a 1.

Indeed, to Turing’s fundamental theorem in computability theory that the halting problem is unsolvable, there corresponds in algorithmic information theory my theorem [4] that the halting probability Ω is a random real number. In other words, any program that calculates N bits of the binary expansion of Ω is no better than a table look-up, because it must itself be at least N bits long. I.e., Ω is incompressible, irreducible information.

And it is Ω itself that is our abstract example of evolution! For even though Ω is of infinite complexity, it is the limit of a computable sequence of rational numbers, each of which is of finite but eventually increasing complexity. Here of course I am using the word “complexity” in the technical sense of algorithmic information theory, in which the complexity of something is measured by the size in bits of the smallest program for calculating it. However this computable sequence of rational numbers converges to Ω very, very slowly.

In precisely what sense are we getting infinite complexity in the limit of infinite time?

Well, it is trivial that in any infinite set of objects, almost all of them are arbitrarily complex, because there are only finitely many objects of bounded complexity. (In fact, there are less than 2^N objects of complexity less than N .) So we should not look at the complexity of each of the rational numbers in the computable sequence that gives Ω in the limit.

The right way to see the complexity increase is to focus on the first K bits of each of the rational numbers in the computable sequence. The complexity of this sequence of K bits initially jumps about but will eventually stay above K .

What precisely is the origin of this metaphor for evolution? Where does this computable sequence of approximations to Ω come from? It arises quite naturally, as I explain in my 1988 *Scientific American* article [13].

The N th approximation to Ω , that is to say, the N th stage in the computable evolution leading in the infinite limit to the violently uncomputable infinitely complex number Ω , is determined as follows. One merely considers all programs up to N bits in size and runs each member of this finite set of programs for N seconds on the standard universal Turing machine. Each program K bits long that halts before its time runs out contributes measure 2^{-K} to the halting probability Ω . Indeed, this is a computable monotone increasing sequence of lower bounds on the value of Ω that converges to Ω , but very, very slowly indeed.

This “evolutionary” model for computing Ω shows that one way to produce algorithmic information or complexity is by doing immense amounts of computation. Indeed, biology has been “computing” using molecular-size components in parallel across the entire surface of the earth for several billion years, which is an awful lot of computing.

On the other hand, an easier way to produce algorithmic information or complexity is, as we have seen, to simply toss a coin. This would seem to be the predominant biological source of algorithmic information, the frozen accidents of the evolutionary trail of mutations that are preserved in DNA.

So two different sources of algorithmic information do seem biologically plausible, and would seem to give rise to different kinds of

algorithmic information.

6. Technical note: A finite version of this model

There is also a “finite” version of this abstract model of evolution. In it one fixes N and constructs a computable infinite sequence $s_t = s(t)$ of N -bit strings, with the property that for all sufficiently large times t , $s_t = s_{t+1}$ is a fixed random N -bit string, i.e., one for which its program-size complexity $H(s_t)$ is not less than its size in bits N . In fact, we can take s_t to be the first N -bit string that cannot be produced by any program less than N bits in size in less than t seconds.

In a sense, the N bits of information in s_t for t large are coming from t itself. So one way to state this, is that knowing a sufficiently large natural number t is “equivalent to having an oracle for the halting problem” (as a logician would put it). That is to say, it provides as much information as one wants.

By the way, computations in the limit are extensively discussed in my two papers [5–6], but in connection with questions of interest in algorithmic information theory rather than in biology.

7. Conclusion

To conclude, I must emphasize a number of disclaimers.

First of all, Ω is a metaphor for evolution only in an extremely abstract mathematical sense. The measures of complexity that I use, while very pretty mathematically, pay for this prettiness by having limited contact with the real world.

In particular, I postulate no limit on the amount of time that may be taken to compute an object from its minimal-size description, as long as the amount of time is finite. Nine months is already a long time to ask a woman to devote to producing a working human infant from its DNA description. A pregnancy of a billion years, while okay in algorithmic information theory, is ridiculous in a biological context.

Yet I think it would also be a mistake to underestimate the significance of these steps in the direction of a fundamental mathematical theory of evolution. For it is important to start bringing rigorous concepts and mathematical proofs into the discussion of these absolutely fundamental biological questions, and this, although to a very limited extent, has been achieved.

References

Items 1 to 10 are reprinted in item 12.

- [1] G. J. Chaitin, “To a mathematical definition of ‘life’,” *ACM SICTACT News*, January 1970, pp. 12–18.
- [2] G. J. Chaitin, “Information-theoretic computational complexity,” *IEEE Transactions on Information Theory* IT-20 (1974), pp. 10–15.
- [3] G. J. Chaitin, “Randomness and mathematical proof,” *Scientific American*, May 1975, pp. 47–52.
- [4] G. J. Chaitin, “A theory of program size formally identical to information theory,” *Journal of the ACM* 22 (1975), pp. 329–340.
- [5] G. J. Chaitin, “Algorithmic entropy of sets,” *Computers & Mathematics with Applications* 2 (1976), pp. 233–245.
- [6] G. J. Chaitin, “Program size, oracles, and the jump operation,” *Osaka Journal of Mathematics* 14 (1977), pp. 139–149.
- [7] G. J. Chaitin, “Algorithmic information theory,” *IBM Journal of Research and Development* 21 (1977), pp. 350–359, 496.
- [8] G. J. Chaitin, “Toward a mathematical definition of ‘life’,” in R.D. Levine and M. Tribus, *The Maximum Entropy Formalism*, MIT Press, 1979, pp. 477–498.
- [9] G. J. Chaitin, “Algorithmic information theory,” in *Encyclopedia of Statistical Sciences*, Volume 1, Wiley, 1982, pp. 38–41.

- [10] G. J. Chaitin, “Gödel’s theorem and information,” *International Journal of Theoretical Physics* 22 (1982), pp. 941–954.
- [11] G. J. Chaitin, *Algorithmic Information Theory*, Cambridge University Press, 1987.
- [12] G. J. Chaitin, *Information, Randomness & Incompleteness—Papers on Algorithmic Information Theory*, World Scientific, 1987.
- [13] G. J. Chaitin, “Randomness in arithmetic,” *Scientific American*, July 1988, pp. 80–85.
- [14] P. Davies, “A new science of complexity,” *New Scientist*, 26 November 1988, pp. 48–50.
- [15] J. P. Delahaye, “Une extension spectaculaire du théorème de Gödel: l’équation de Chaitin,” *La Recherche*, juin 1988, pp. 860–862. English translation, *AMS Notices*, October 1989, pp. 984–987.
- [16] I. Stewart, “The ultimate in undecidability,” *Nature*, 10 March 1988, pp. 115–116.