Order, Organization, and Randomness: On the Mathematical Formulation of Life

The idea that life is to be understood in terms of information has strongly taken hold in recent decades. I discuss two attempts to carry this through mathematically. G. J. Chaitin, co-founder of algorithmic information theory, proposes an information-theoretic definition of life in terms of organized complexity (Chaitin 1990a and 1990b).[[1]](#footnote-1) More recently, William Dembski, Winston Ewart, and Robert Marks have attempted to formulate in information-theoretic terms Dembski’s concept of specified complexity, using a mathematically hybrid entity they term “algorithmic specified complexity” (Ewart, Dembski, and Marks 2013a, 2014, 2015a, 2015b), and Dembski and Ewart have reformulated this concept in their newly revised edition (Dembski and Ewart 2023) of Dembski’s *The Design Inference* (Dembski 1998).[[2]](#footnote-2) The aim in both cases is to mathematically distinguish informational properties of biological complexity, in contrast to simple order, on the one hand, and mere randomness on the other. Moreover, the respective mathematical strategies are the same: To take an informational measure and subtract out its randomness, leaving a remainder of organization (Chaitin) or specified complexity (Dembski et al.).

The purpose of this paper is to show, using the attempts of Chaitin and Dembski et al. as test cases, that in principle biological complexity cannot be defined or detected in information-theoretic terms. More specifically, I shall argue that while information theory can distinguish order from randomness and thus detect ordered specifications—cheating in a card game, for instance—any attempt to detect biological organization in information-theoretic terms must founder on the information-theoretic indistinguishability of organization and randomness. That is, such methods can indirectly detect combinations of order and randomness, but not organization in the sense of life.[[3]](#footnote-3) The attempts of Chaitin and Ewart et al. fail in an instructive way, however, by illustrating the inability of mathematical methods generally to register meaning or organizational structure and so either to detect biological complexity or formulate a definition of life.[[4]](#footnote-4)

1. Chaitin: Algorithmic Formulation of Organization

While Chaitin makes no claim to provide a definitive account of life in information-theoretic terms, he does aim to at least set forth tentatively “a rigorous quantitative definition” of concepts like “life” and “organism” (Chaitin 1990b, 92). Chaitin grounds his attempt in a proposed informational measure of organization.[[5]](#footnote-5) His governing idea is that a living organism, taken as a unity in contradistinction to a mere aggregation of parts, should be computable, as it were, by means of a program substantially smaller than a program exhaustively computing the assemblage of each and every part:

A living being is a unity; it is simpler to view a living being as a whole than as a sum of its parts. If we want to compute a complete description of the region of space-time that is a living being, the program will be smaller in size if the calculation is done altogether, than if it is done by independently calculating descriptions of parts of the region and then putting them together. (Chaitin 1990a, 85)

That is to say, unlike either simple order (a perfect crystal, for example) or randomness (an ideal gas), life is *organized*. By analogy to a geometrical structure, biological organization in Chaitin’s formulation is to be measured by the increasing difference in bits between the algorithmically compressed or minimal “holistic” program and the minimal “aggregate program” , as the maximum size of parts included in the aggregate program becomes progressively smaller, yielding a more and more fine-grained description (Chaitin 1990b, 93-94). The decisive feature of Chaitin’s prescription for is that “[e]ach piece must have a separate description that makes no cross-references to any of the others” (Chaitin 1990b, 94), such that descriptions previously compressible in the holistic program are no longer compressible. On Chaitin’s account, then, as inflates with decreasing , the degree of organization is given by . For instance, the assembly of a pair of human lungs (my example) would be computable in our holistic program by taking the minimal program describing a lung cell and adding the instruction, “Put this many of them together like so to make a lung and then put another lung over here,” whereas once has decreased below the diameter of an individual lung cell reference to other lung cells is proscribed and so the aggregate program must compute the disposition of each lung cell individually. This yields a much longer program such that is maximized. And since a minimal program measures complexity, Chaitin’s procedure, he maintains, registers organization in terms of the increase in complexity as the description of a living organism becomes more and more fine-grained:

Hence as a function of is a kind of ‘spectrum’ or ‘Fourier transform’ of . will increase as decreases past the diameter of significant patterns in , and if is organized hierarchically this will happen at each level in the hierarchy.

Thus the faster the difference increases between and as decreases, the more interrelated, structured, and organized is. (Chaitin 1990b, 94)

By contrast, for both randomness and simple order cannot decrease past the diameter of significant patterns, either because there are no such patterns (randomness) or because such patterns go all the way down, as it were (simple order).

The reference to an increase in the difference between and “at each level in the hierarchy” suggests Chaitin has in mind a stepwise increase in that difference as one moves from one level of the hierarchy down to the next. Yet Chaitin also seems to suggest that the decisive property for organization is rather the *rate of change* of the difference between and as decreases. These two criteria, though, are distinct from one another and it is not clear how they cohere in Chaitin’s account.

A deeper problem with Chaitin’s analysis, though, is that organization is indistinguishable in information-theoretic terms from randomness: both are high complexity and therefore high information, low order, and low algorithmic compressibility. But organization in the sense of life distinguishes itself by the mutual adaptation of parts (organs) for the sake of the organism as a whole. Thus Kant in his *Critique of Judgment* (1790) defines an “organized being” in the sense of biological life as a “product of nature . . . in which everything [each part] is a purpose and reciprocally a means” (Kant 1987, 255).[[6]](#footnote-6) The human hand, for instance, which in serving the interests of the whole organism serves the heart, is in turn served by the heart, and therefore is both means and end (purpose). Such mutual adaptation of parts is definitive of organization not just as pertains to life but more generally (meaningful text, machines, music, and so forth). Organization taken this way is in no wise the same as order in the sense of compressibility: for an inherent feature of organized beings is high complexity, such that their specification requires a long and comparatively incompressible algorithm. Algorithmic information theory distinguishes order from randomness, but it does not distinguish organization from randomness.

Chaitin’s formula is equivalent conceptually to the information-theoretic concept of *randomness deficiency*,or the difference between the uncompressed length of a bit string and the minimal program computing that string.[[7]](#footnote-7) While a random sequence by definition cannot be compressed, an ordered sequence can—for instance, a one hundred bit alternating sequence of ones and zeros can be compressed as, “Print 10 and repeat 50 times,” and an alternating sequence of ones and zeros a million bits long can be compressed as, “Print 10 and repeat 500,000 times.” Thus, while both sequences are non-random, the million-bit string has far more randomness deficiency, and therefore also has “more order,” if you will, than the hundred-bit string (and on that account is less likely to have occurred by chance[[8]](#footnote-8)). Chaitin’s essentially gives the randomness deficiency for progressively smaller values of since his proscription on cross-references to other parts in the term renders it progressively uncompressed. And since is the minimal (maximally compressed) program, Chaitin’s formula amounts conceptually to a randomness deficiency. But a deficiency of randomness is not tantamount to organization in the sense of life. A string consisting of all 1’s, for instance—Chaitin’s “crystal”—will exhibit the highest rate of increase in randomness deficiency as decreases—not because the string is organized but because it is maximally ordered and therefore maximally compressible. Chaitin knows this better than anyone, of course (“Neither a gas nor a crystal should count as organized . . .” [Chaitin 1990b, 93]), but in the course of his analysis he often seems to forget it.

With respect to the detection of hierarchical structure, Chaitin offers for illustration a sequence of bits, of length , in which starting with 1 the string is concatenated with itself by progressive doubling, according to the rule that at the th stage the sequence is either repeated or replaced with its complement depending on whether the th bit is a 0 or 1 respectively (Chaitin 1990b, 104–105). This yields a random sequence of complements and repetitions , with each consecutive pair of segments compressible as ( = length of the combined segments). It follows that for large , . And so as decreases, increases according to (where ), and so the difference  increases.[[9]](#footnote-9) Thus the descent through each level of the hierarchy is accompanied by a step change in as decreases below the level of significant patterns in .

Chaitin’s formula has registered a hierarchically ordered arrangement of internally complex parts, and such hierarchical structure indeed is characteristic of life. A human lung cell, once again, is algorithmically complex while the symmetrical arrangement of two lungs is algorithmically simple. Thus the human body exhibits hierarchical structure in the sense of a comparatively large value of . If we came upon Chaitin’s hierarchically concatenated bit string above, we would be hard-pressed to explain the repetitions and complements of lengthy complex segments either by chance or natural necessity. But we would not have detected biological organization, for the hierarchically ordered bit strings in Chaitin’s scenario above are internally random. Chaitin thus is imprecise when, in the passage quoted above, he speaks of life being “interrelated, structured, and organized,” as if these three meant the same thing. Organization as it pertains to life is not equivalent to mere structure or interrelatedness, and since organization, in the sense of high information phenomena such as life, machines, natural language, and so forth, is not algorithmically distinct from randomness, what Chaitin’s method really detects is the symmetrical arrangement of algorithmically random parts—that is, parts describable internally by an algorithmically complex program. These algorithmically complex parts themselves may be either organized or merely random, and algorithmic information theory is blind to the difference. The symmetrical branches of a snowflake, for instance, are internally complex but not internally organized.

A more promising suggestion on Chaitin’s part, at first blush at least, is to measure organized structure in terms of mutual information[[10]](#footnote-10):

The information content of an organism contains much irrelevant detail, and a bigger animal is necessarily more complex in this sense. *But if it were possible to calculate the mutual information of two arbitrary cells in a body at a given moment, we surmise that this would give a measure of the genetic information in a cell. This is because the irrelevant details in each of them, such as the exact position and velocity of each molecule, are uncorrelated and would cancel each other out*. (Chaitin 1990b, 94)*.*

The idea that organization should distinguish itself by high mutual information between complex parts is plausible initially.[[11]](#footnote-11) Our two lungs, for instance, have high mutual information because the organized complexity of each is repeated in the other. It seems unlikely that two independent random sequences should exhibit high mutual information. It is, to be sure, unlikely by chance, but not so unlikely by natural necessity. A snowflake, once again, exhibits high mutual information between its symmetrical branches, but “no two are alike” because the branches themselves are internally complex and therefore algorithmically random. Thus the snowflake exhibits merely a combination of order and randomness—not organization in the relevant sense. Similarly, the replication of a sequence of so-called “junk DNA” would produce two randomly complex and therefore unorganized strands with high mutual information.

Chaitin advances a number of theorems which, he maintains, quantify various degrees of organization (“gas,” “crystal,” “bilateral symmetry,” “hierarchy”), and one can hardly doubt the proofs (Chaitin 1990b, 102-105). But the problem is philosophical rather than mathematical: For even if life for the most part does incorporate symmetry, life is not defined by symmetry—even hierarchically structured symmetry. That is why conceptually we must distinguish organization (an arrangement of *organs*) from mere order (pattern or algorithmic compressibility). If a sequence of DNA bases, once again, were to exhibit the structure of Chaitin’s “bilateral symmetry” we surely would conclude it did not come from a living organism, for a highly compressible sequence of DNA bases would not carry sufficient information to specify a living organism. Organization, once again, is neither a combination of, nor compromise between, order and randomness. In general, since algorithmic complexity is equivalent to randomness, we cannot subtract out randomness without subtracting out organized complexity along with it. Thus, while Chaitin’s formula can register hierarchy, it cannot define life.

2. Dembski, Ewart, and Marks: Algorithmic Specified Complexity

William Dembski and his co-authors Ewart and Marks attempt to formulate in information-theoretic terms the notion of *specified complexity*, advanced most notably by Dembski in connection with his well-known “design inference.”[[12]](#footnote-12) While the authors evidently are unaware of Chaitin’s earlier papers, the logic of algorithmic specified complexity (ASC) parallels Chaitin’s: subtract out the randomness from a longer bit string to secure a remainder of organized complexity.[[13]](#footnote-13) Here we first review the logic of Dembski’s design inference, then the concept of conditional Kolmogorov complexity appropriated by the authors as “contextual information,” and finally the concept of algorithmic specified complexity itself.

2.1 *Dembski’s Explanatory Filter*. Dembski’s concept of specified complexity is best presented through a flow chart he dubs the “explanatory filter” (see, in particular, *The Design Inference* [1998, 2023] and *No Free Lunch* [2002]).[[14]](#footnote-14) Entering the filter, we ask first whether the phenomenon or event under investigation is *contingent* in the sense of being consistent with, while not determined by, natural regularities. For example, a roll of the dice conforms to the laws of physics (gravity, Newtonian force, and so forth) but still leaves room for a variety of outcomes. Similarly, the deal of a shuffled deck is consistent with natural regularities but does not dictate a specific sequence of cards. Let us then take as a contingent event the deal of three consecutive royal flushes in a poker game. Proceeding to the next stage of the explanatory filter, we ask whether the event in question is complex in the sense of low probability. (“Complexity” in Dembski’s sense corresponds to low probability, or high Shannon information, and should not be confused with algorithmic complexity.) Three consecutive royal flushes (probability 1 in ) is far less probable than, for example, three consecutive pairs (about 1 in 13), and so the sequence of royal flushes is far more complex.[[15]](#footnote-15)

Finally, given a contingent event of high complexity, we ask whether the event is *specified* in the sense of exhibiting an independently identifiable pattern. In our case, the royal flushes indeed exhibit such a pattern—five cards of the same suit in sequential order, predesignated as the highest-ranked hand. Thus, we can classify three consecutive royal flushes as an event exhibiting high specified complexity and we are justified in suspecting intelligent design (cheating).

While Dembski’s various characterizations of specified complexity have come in for criticism by biologists, mathematicians, and philosophers, the idea itself is not original with Dembski and his account strikes me as clear enough in general, even if a formal definition is somewhat elusive.[[16]](#footnote-16) Among obvious examples of specified complexity—besides dishonest poker games—are, for example, written text, the DNA sequences of living organisms, musical compositions, and so forth. We know specified complexity when we see it. Nevertheless, Dembski’s use of “pattern” to define the concept of specification is unfortunate since the term suggests periodicity or repetition (a striped shirt, for instance, or the design of a chessboard) and not all specifications are patterned in this sense. A DNA sequence, once again, is not patterned and indeed could not carry appreciable genetic information if it were. In the context of algorithmic information theory, patterned sequences fall into the class of highly compressible and thus low complexity and low information sequences.[[17]](#footnote-17) Thus the presence of a pattern, at least in the sense of periodicity, is not a necessary condition for specification since organized specifications exhibit high algorithmic complexity and therefore low order. Neither is it a sufficient condition since pattern can arise through natural necessity (a chemical crystal or snowflake, for instance). Moreover, even a random sequence can be specified, as, for example, dealing a random poker hand at the beginning of the evening and then designating it the highest ranked hand for the rest of the night.[[18]](#footnote-18) It is, in the end, difficult to define “specification” apart from the idea of intelligent design, for specification suggests a *specifier*. Perhaps “organized complexity” would be more neutral in this respect. For our purposes, at any rate, a specification in general may be regarded as something singled out from all the possibilities, such that it appears to have been designated.[[19]](#footnote-19)

2.2 *Conditional Complexity*. Shannon information (or Shannon “entropy”), developed in the context of communication theory, is a measure of the uncertainty resolved when a message is received, as per the equation for a binary code ( designating the information per digit in the received sequence and the respective probabilities of the symbols 0 and 1). We define as one bit (binary digit) of information for a binary code of equiprobable symbols, and so the total information in such a sequence is simply its length . In terms of probability, the total information in a sequence of probability is , since the number of possible arrangements of the symbols is . Thus, although there exists no measure of complexity per se in Shannon information theory, by defining complexity in terms of improbability Dembski can employ Shannon information as a measure of complexity.

The concept of *conditional entropy* is introduced by Shannon in the context of measuring information loss in a communications channel due to transmission error (noise). Suppose characters in a binary transmission are erroneous at the rate of 1%. One might be tempted, Shannon notes, to conclude that 1% of the information at the source has been lost, but that is not the case since we do not know which of the received characters are the erroneous ones (a 50% error rate in a binary transmission, after all, would result in a 100% information loss). Since Shannon information measures how much uncertainty is resolved when a message is received, the correct measure of information loss is the remaining uncertainty of what was transmitted or the entropy of the source bit string given the received bit string : , where is the probability of the joint occurrence of the two events and at the source and receiver respectively, and is the uncertainty (entropy) of the joint event. That is, for pairs of events and , to determine the information loss we take the weighted average of the uncertainty of events at the source given our knowledge of events at the receiver (Shannon 1948, 407–408). Abstracted from the context of noise, conditional entropy measures the amount of additional information we would need to derive a sequence , given our knowledge of a sequence .

Conditional Kolomogorov complexity, proposed by Andrei Kolmogorov in 1968 as a “new definition” of Shannon conditional entropy, is designed to attach a “definite meaning” to the concept of conditional entropy without recourse to probability (Kolmogorov 1968, 662). Kolmogorov interprets Shannon conditional entropy algorithmically as the “minimal length of the recorded sequence of zeros and ones of a ‘program’ *P* that permits construction of the value of , the value of being known” (Kolmogorov 1968, 662). Thus, conditional Kolmogorov complexity is the algorithmically compressed equivalent of Shannon conditional entropy . And since is a minimal program it is also a *random sequence* as defined in algorithmic information theory.

2.3 *Algorithmic Specified Complexity*.[[20]](#footnote-20) Dembski and his coauthors note that neither Shannon information nor algorithmic complexity registers *meaning* (Ewart et al. 2014, 132-133). As Shannon remarks famously in the introduction to his 1948 paper, it is only “frequently . . . [that] messages have *meaning*” (Shannon 1948, 379). Kolmogorov complexity , for its part, measures not meaning but randomness. Accordingly, Dembski et al. suggest, to detect meaningful information we might take the Shannon information (complexity) of a sequence and subtract out the amount of randomness: “Since high compressibility corresponds to specification, the compressed length of the string is subtracted. Thus, high improbability counts for specified complexity, but incompressible strings count against it” (Ewart Et al. 2014, 136). This analysis yields for algorithmic specified complexity, where is the Shannon information, is the contextual information, and is the conditional Kolmogorov complexity.

We note once again, as in Chaitin’s account, a conceptual relation to the randomness deficiency in algorithmic information theory—the difference between the length of a program and its compressed or minimal program.[[21]](#footnote-21) However, as we observed already in our discussion of Chaitin’s proposal, randomness deficiency cannot itself register organization, but tells us merely where we are on the continuum between order and randomness. The innovation proposed by Dembski et al. is to register specified complexity through what they call “contextual information.” In accordance with this line of thought, the conditional Kolmogorov complexity functions as a measure of the algorithmic information contained by the sequence under investigation, given our knowledge of contextual information . This contextual information, which pertains to specification, thus allows for the algorithmic compression of a complex specified sequence.

Let us take, for instance, our poker cheating scenario. Consider the contextual information to be a list of ranked poker hands. Such a list registers meaning since only designated patterns count as specifications in poker, not all patterns (a poker hand of all cards of the same color, for instance, does not count as a specification, while a flush, consisting of all cards of the same suit, does). Assuming the deal is honest and therefore random, the contextual information , in the form of a list of ranked hands, fails to supply us any information that would help us derive the sequence actually dealt. Thus, and (or close to zero). We thereby establish a benchmark for the absence of specified complexity. Given a cheating scenario, however, the background information (list of ranked hands) allows us to algorithmically compress the dealt sequence. Suppose I repeatedly deal myself four aces in a game of five card draw. This sequence of poker hands, given the contextual information, is encodable as a highly compressible bit string. By contrast, the sequence of random hands in an honest poker game is comparatively incompressible because it must be listed card by card. For the cheating scenario, accordingly, the conditional Kolmogorov complexity is much lower while the complexity is the same as for the random scenario, yielding a higher value of ASC. And, since all else equal most sequences are random and therefore incompressible, a sequence of poker hands exhibiting high compressibility makes it more likely that cheating is in play. So algorithmic specified complexity in this case appears to succeed in detecting intelligent design.

However, the same result would hold even without the contextual information concerning the ranking of poker hands since algorithmic compressibility detects ordered patterns in general. It is the high randomness deficiency that signals cheating, not algorithmic specified complexity per se. Indeed, the idea of “contextual information” can only confuse the issue, for the randomness of a sequence, according to algorithmic information theory, is a *characteristic of the sequence itself*, apart from any knowledge of its origin or other background information. Thus, the only allowable contextual information pertains to the list of symbols employed to encode the information. In a standard binary sequence, ones and zeros count as symbols and any other characteristics of the symbols beyond that (size, color, or the like) are irrelevant. For our poker scenario, suit and numerical rank define our symbols whereas both color and hand rankings are irrelevant. Given a system of symbolic encoding, then, no additional contextual information is relevant to an information-theoretic analysis. The introduction of conditional Kolmogorov complexity into the ASC measure therefore is logically incoherent.

As the authors note, the specifications of greatest interest are rarely simple patterns (Ewart et al. 2014, 135), but rather complex ones such as are exhibited by living organisms, meaningful text, and so forth. Unfortunately, the authors do not always keep this distinction in mind, for they consistently treat specified complexity as the opposite of randomness, whereas in algorithmic information theory randomness and complexity are equivalent informational measures. Let us consider the example the authors provide for the case of natural language text (Ewart et al. 2014, 137-140). The authors here propose to detect specified complexity in the sentence, “The quick brown fox jumps over the lazy dog.” Encoding the sentence in UTF-32, a 32 bit per character code, yields a total of 1,376 bits of Shannon information for . The authors note further that given contextual information consisting of knowledge of the English alphabet (26 letters plus the space) the sentence can be compressed into about 205 bits.[[22]](#footnote-22) Substituting 205 into the equation as the conditional Kolmogorov complexity yields or about bits of algorithmic specified complexity. Were the bit string under investigation random, by contrast, containing no meaningful information, then regardless of the contextual information, the authors suggest, the string would be incompressible: would equal and the algorithmic specified complexity would be zero. The authors accordingly conclude that their ASC measure has detected specified complexity since “the large number of bits gives a good indication that it is highly unlikely that this sentence was generated by randomly choosing bits” (Ewart Et al. 2014, 138).

This conclusion does not follow. The compressibility of the UTF-32 encoded sentence once again has nothing to do with background information. UTF-32 employs a far larger number of bits per character than is required to encode the sentence under consideration. Therefore, most of the bit string consists of leading zeros. It is the compressibility of these leading zeros that is registered by the low Kolmogorov complexity , apart from knowledge of the contextual information, and therefore entirely apart from . Once again, we detect the compressibility of the UTF-32 sequence by its randomness deficiency , not its algorithmic specified complexity . This is obvious if we imagine, in place of the sentence of interest, a random sequence of 43 English letters encoded in UTF-32. For the UTF-32 case would still be about 1,376 and for the random case, without UTF-32 coding, still about 205, yielding the same high value for algorithmic specified complexity. But in fact, the only thing that has been detected by the ASC measure is the inefficiency of UTF-32 for encoding the English alphabet.[[23]](#footnote-23) Alternatively, we could regard the same scenario in terms of optimal Shannon coding. UTF-32 clearly is suboptimal for encoding the English alphabet since we are employing a 32-bit per character code that could be compressed to 5 bits per character if, for instance, we set 00000000000000000000000000010011=10011 (for some letter).[[24]](#footnote-24)

The authors overlook a different kind of analysis which might have enabled the algorithmic detection of meaningful text in contrast to a random sequence of letters. Meaningful text contains redundancies that allow for some degree of algorithmic compression. For example, the letters in meaningful English text are not equiprobable: the letter “e,” for instance, occurs far more often than “q,” the letter “u” almost always following the letter “q,” and so forth. [[25]](#footnote-25) Consequently, meaningful text is somewhat compressible algorithmically as opposed to a purely random sequence of letters, but still far less compressible than an ordered sequence.[[26]](#footnote-26) In terms of Shannon optimal coding we might, for example, set . In this case, then, ASC again reduces to the randomness deficiency or the difference between the respective lengths of the compressed and uncompressed descriptions. But, to repeat, the randomness deficiency cannot distinguish organization from randomness. What is detected by means of a randomness deficiency is order, not organization. To be sure, absent a plausible natural mechanism to produce such order we infer that conventional redundancies are in play. But that is not an information-theoretic detection of organization.

To further bring home this point, observe that were an organism’s DNA written out for examination without our knowing beforehand the sequence pertained to a living being, we would not find much to distinguish it in information-theoretic terms from a random sequence of DNA bases.[[27]](#footnote-27) The only qualification is that, like meaningful English text, DNA contains redundancies that allow for some degree of algorithmic compression (for instance, not all DNA bases are equiprobable), thereby reducing its algorithmic complexity.[[28]](#footnote-28) Schmitt and Herzel (1997) find, however, that real DNA sequences are closer to completely random sequences than to written text.[[29]](#footnote-29)

Nevertheless, functional DNA is similar to meaningful text in that it exhibits a degree of randomness deficiency. The difference, however, is that in the case of text we start with a set of symbols (alphabet) known to have been invented by human beings: we infer meaning because there is no plausible natural process by which a partially compressible or redundant sequence of letters of the alphabet could arise. And since a chance process almost always generates a random sequence, a partially compressible sequence of letters signals specification and, in fact, deliberate intent. But there are conceivable natural processes that could generate a partially compressible sequence of DNA bases and thus distinguish that sequence from a purely random sequence, without the DNA sequence being specified. Imagine a sequence of DNA bases formed by chance in the proverbial “primordial soup,” where there would not have been equal availability of the four DNA bases. Our chance sequence would exhibit some degree of algorithmic compressibility and thus distinguish itself from an entirely random sequence of equiprobable DNA bases. But it would not necessarily distinguish itself from a functional sequence, which also is partially compressible. So we could not distinguish biological complexity from random complexity based solely on the randomness deficiency. We would need, for instance, a prior knowledge of the expected natural availability of DNA bases versus their probabilities in functional DNA sequences. But in that case we are not detecting functional DNA or life in strictly information-theoretic terms. The same must apply to the other examples Dembski et al. suggest demonstrate the efficacy of algorithmic specified complexity for detecting biological complexity.[[30]](#footnote-30)

2.4 *Specification and Minimum Descriptive Length*

In their revised edition of Dembski’s *The Design Inference* (1998), Dembski and Ewart argue that specification is definable algorithmically in terms of a short description:[[31]](#footnote-31)

We call patterns that have a simple description specifications. But what makes a description simple? To answer this question with mathematical precision requires measuring the length of descriptions and then designating the simple descriptions as those that have short length. This approach turns specification into a complexity-theoretic notion, making descriptions more or less complex depending on their length. The underlying metric that measures description-length is therefore a complexity measure. (Dembski and Ewart 2023, 134)

The previous strategy for measuring specified complexity remains in effect, though: to detect specified complexity by combing high probabilistic complexity (that is, low probability) with low algorithmic complexity (high order): “Design inferences engage in a balancing act between the complexity of events and the simplicity of descriptions” (133). The main difference in the revised account is that the specificity of an event is no longer given in terms of conditional Kolmogorov complexity, but rather more generally in terms of the length of the shortest possible description of the event under consideration: , where is the complexity measured in terms of improbability or Shannon information and is simply the minimum description length. A royal flush in poker, for instance, exhibits specified complexity because it has high probabilistic complexity (1 in 649,740 chance) along with a short description (simply the phrase “royal flush”), whereas any hand whatever, while likewise having a short description (“any hand”) has a probability of 1 (Dembski and Ewart 2023, 139). Similarly, according to the authors, Lincoln’s “Gettysburg Address” has high complexity (with 29 different characters and a length of approximately 1,500 characters, the probability of the text is for a total of or 7,286 bits of complexity) while it is specifiable by the short description “Gettysburg Address” (about 40 bits assuming 20 bits per word times two words). Lincoln’s text, then, has bits of specified complexity (Dembski and Ewart 2023, 301–302).

The authors’ revised formulation, however, is plagued by the same conceptual flaws as the original formulation of ASC discussed above (2.3). In the first place, in terms of algorithmic information theory, short descriptions are characteristic of ordered, low information specifications. Moreover, descriptions such as “royal flush” and “Gettysburg Address” are not compressions in the sense of algorithmic information theory, for, as we have noted, algorithmic compressibility is a feature of a sequence itself, apart from background information. The royal flush, to be sure, is compressible algorithmically in terms of the actual symbols on the cards (suit and numerical rank), but it is not compressed by the phrase “royal flush,” which rather is a name, not a minimal description in the algorithmic sense. And Lincoln’s text is not compressible except for the redundancies of the English language. The title “Gettysburg Address” is not an algorithmic compression of 7,286 bits of Shannon information, but rather the replacement of those 7,286 bits with the name of the speech. We might as well single out a particular random sequence and shorten its description by naming it “Random Sequence #1,” yielding high ASC. No doubt if Random Sequence #1 comes up again we will infer intelligent design, but that has nothing to do with compressibility in the sense of algorithmic information theory. The algorithmic compression of a complex sequence is a contradiction in terms.

Dembski and Ewart offer more promising illustration of algorithmic specified complexity in the Voynich Manuscript, a 240-page document in unknown fifteenth-century handwritten script (Dembski and Ewart 2023, 305–307). Here the authors’ appeal to algorithmic compressibility is legitimately based on non-random frequency of characters and words—that is, textual redundancy as discussed above. Taking the length of the Voynich manuscript as approximately 30,000 words, five percent of which are five particular words of five characters each, Dembski and Ewart calculate the improbability of the sequence as about 21,749 bits of complexity and—based on 12 words at 20 bits per word for the minimum description “five words of five characters make up five percent of the text”—estimate the description length as 240 bits of specification.[[32]](#footnote-32) Thus the authors calculate or about 21,509 bits of specified complexity. But the calculation is incorrect since 240 bits is the minimum description length of the five percent only, whereas the correct measure of is the minimum or maximally compressed description length of the entire text: not 240 but about . The true measure of algorithmic specified complexity thus is or 847 bits, in line with the modest randomness deficiency we would expect for partially compressible (due to redundancy) written text. This modest randomness deficiency does detect a degree of order inexplicable in terms of natural mechanisms, but once again a subsequent inference to organization (meaning) goes beyond an information-theoretic analysis.

3. Conclusion: Order, Organization, and Randomness

Life is organized, not ordered per se, although life typically is ordered in some respects (body plan symmetries, for instance). But organization, biological or otherwise, is in information-theoretic terms indistinguishable from randomness. Thus, only ordered or patterned specifications can be detected by the information-theoretic approach, not organized specifications. And solely because order is unlikely to arise by chance can we infer design in such cases. With respect to the debate over the intelligent design of life, the quest for a mathematical definition of specified complexity or biological organization is irrelevant. We can use information theory to estimate how much meaningful information, for example, would have been required for the origin of life on earth (at least about 700,000–800,000 bits by best estimate[[33]](#footnote-33)) or for some specific evolutionary increase in biological complexity (like the Cambrian Explosion), but the question of intelligent design hinges solely on whether that amount of meaningful information could have arisen by chance.[[34]](#footnote-34) Organization cannot arise by natural necessity and is unlikely—perhaps impossible—by chance.

Claude Shannon in 1948 famously distinguished abstracted mathematical information from meaningful information, which raises the question, If life cannot be defined in terms of mathematical information, is it definable in terms of meaningful information? The answer, I suggest, is no. Life transcends not just physics and chemistry (so Polanyi), but information as well. A DNA sequence contains meaningful information, but the DNA molecule itself is neither alive nor is it the “secret of life.” DNA does not replicate itself but is replicated by the cell. DNA figures in biological life solely through the role it plays in the living cell and apart from the cell the sequence of DNA bases encodes nothing and carries no biological information.[[35]](#footnote-35) Information encoded in DNA is a necessary condition for biological life as we know it, but what makes the cell is alive in the first place cannot be explained by information. Information theory, to be sure, has a crucial role to play in molecular biology because of the strong analogy between digital information and the DNA encoding system in living cells. But we should avoid inflated claims—such as suggested by overwrought slogans like “It from Bit” and so forth—either for information itself or for information theory. Life explains information, not the other way around.

Declarations

Funding: This work was supported by a Providence College 2021-2022 Interdisciplinary Scholarship Award.

The author has no relevant financial or non-financial interest to disclose.

References

Abel, David L. and Jack T. Trevors (2005). Three subsets of sequence complexity and their

relevance to biopolymeric information. *Theoretical Biology and Medical Modelling* 2:29.

<https://tbiomed.biomedcentral.com/counter/pdf/10.1186/1742-4682-2-29.pdf>,

accessed November 5, 2023.

Adriaans, Peter (2008). Learning and the cooperative computational universe. In *Philosophy of*

*Information*, 133–167. North-Holland.

Chaitin, G. J. (1990). *Information, randomness, and incompleteness: papers on algorithmic*

*information theory.* World Scientific Publishing Co.

\_\_\_\_\_\_\_ (1990a). To a mathematical definition of “life.” In: Chaitin 1990, 85–91. First published in 1970.

\_\_\_\_\_\_\_ (1990b) Toward a mathematical definition of “life.” In: Chaitin 1990, 92–110. First

published in 1979.

Dalkilic, Mehmet, Wyatt T. Clark, James C. Costello, and Predrag Radivojak (2006). Using

Compression to Identify Classes of Inauthentic Texts. *Proceedings of the 2006 SIAM*

*Conference on Data Mining.*

[*http://www.khoury.northeastern.edu/home/radivojac/papers/dalkilic\_sdm\_2006.pdf*](http://www.khoury.northeastern.edu/home/radivojac/papers/dalkilic_sdm_2006.pdf),

accessed November 3, 2023.

Dembski, William (1998). *The design inference: eliminating chance through small probabilities.*

Cambridge.

\_\_\_\_\_\_\_ (2002). *No free lunch: why specified complexity cannot be purchased without*

*intelligence.* Rowman & Littlefield.

\_\_\_\_\_\_\_ (2005). Specification: the pattern that signifies intelligence. *Philosophia Christi* 7:

299–343.

Dembski, William and Douglas Ewart (2023). *The Design Inference*, 2nd edition, revised and

expanded. Seattle: Discovery Institute Press.

Elsberry, Wesley and Jeffrey Shallit (2011). Information theory, evolutionary computation, and

Dembski’s “complex specified information.” *Synthese* (2011): 237–270.

Ewart, Winston, William Dembski, and Robert J. Marks II (2013a). On the improbability of

algorithmic specified complexity. In *45th Southeastern Symposium on Systems Theory,*

*68–70. IEEE*.

\_\_\_\_\_\_\_ (2013b). Active information in metabiology. *BIO-Complexity* 4:1–10.

https://doi:10.5048/BIO-C.2013.4

\_\_\_\_\_\_\_ (2014). Algorithmic specified complexity. In *Engineering and the Ultimate: An*

*Investigation of Order and Design in Nature and Craft*, edited by J. Bartlett, D. Halsmer,

and M. Hall, 131–151. Blythe Institute.

\_\_\_\_\_\_\_ (2015a). Measuring meaningful information in images: algorithmic specified

complexity. *IET Computer Vision* 9:884–894. doi:1049/iet-cvi.2014.0141

\_\_\_\_\_\_\_ (2015b). Algorithmic specified complexity in the game of life. *IEEE Transactions on*

*Systems, Man, and Cybernetics: Systems* 45:584–594.

Fitelson, Branden, Christopher Stephens, and Elliott Sober (1999). How not to detect

design—critical notice: William A. Dembski’s *The Design Inference*. *Philosophy of Science*

66 (September 1999): 472–488.

Kant, Immanuel (1987), *Critique of Judgment*. Translated by Werner S. Pluhar. Indianapolis, IN:

Hackett.

Kolmogorov, Andrei N (1968). Logical basis for information theory and probability theory.

Translated courtesy of AFOSR, USAF. Edited by A. V. Balakrishnan. *IEEE Transactions on*

*Information Theory* 14: 662–664.

Li, Ming and Paul Vitányi (2008). *An introduction to Kolmogorov complexity and its applications*.

Third edition. Springer.

Schmitt, Armin O. and Hanspeter Herzel (1997). Estimating the entropy of DNA sequences.

*Journal of Theoretical Biology* 188: 369–377.

Shannon, Claude (1948). A mathematical theory of communication. *Bell System Technical*

*Journal* 27:379–423 (July) and 623–656 (October).

Yockey, Hubert (1992). *Information theory and molecular biology.* Cambridge.

\_\_\_\_\_\_\_ (2000). Origin of life on earth and Shannon’s theory of communication. *Computers and*

*Chemistry* 24:105–123.

\_\_\_\_\_\_\_ (2005). *Information Theory, Evolution, and the Origin of Life*. Cambridge.

1. I shall not here address Chaitin’s “metabiology,” which proposes an algorithmic model for Darwinian evolution. [↑](#footnote-ref-1)
2. To avoid confusion I shall refer to Dembski and his coauthors Ewart and Marks as “Dembski et al.,” though citations are under Ewart as first author. [↑](#footnote-ref-2)
3. I say “indirectly detect” because information theory abstracts from the internal structure of a data set. I thank a *Synthese* referee for calling my attention to this point. [↑](#footnote-ref-3)
4. I note parenthetically that while Dembski and his coauthors, unlike Chaitin, are supporters of Intelligent Design, the present paper abstains on that question. Its purpose is rather to evaluate the possibility of a mathematical account of life—whatever its origin—in information-theoretic terms. [↑](#footnote-ref-4)
5. Chaitin admits that such a definition would cover anything that is “either alive or is a by-product of living beings and their activities” (Chaitin 1990b, 94-95). Thus, strictly speaking, it is not a definition of life but rather a necessary condition. [↑](#footnote-ref-5)
6. For Kant, as is well known, the notion of natural organization, while indispensable, is only a “regulative idea” to guide scientific investigation, not a constitutive concept, since the concept of matter in natural science excludes purpose. [↑](#footnote-ref-6)
7. Li and Vitányi 2008, 120. [↑](#footnote-ref-7)
8. The vast majority of all possible sequences are random, and so highly ordered sequences are unlikely to occur by chance. [↑](#footnote-ref-8)
9. Tellingly, for large both the string of 1’s and Chaitin’s “hierarchy” scenario yield . [↑](#footnote-ref-9)
10. Mutual information, defined by Claude Shannon in terms of the capacity of a communications channel with noise, is a measure of the information shared by two sequences (Shannon 1948, 407–408). Two independent binary sequences with half their characters in common, for instance, have zero mutual information while two identical sequences have 100% mutual information. The mutual information between two sequences and is defined by Shannon as , where is the conditional entropy or additional information we would need to derive given . Shannon did not himself use the term mutual information. [↑](#footnote-ref-10)
11. I am unable to make sense, however, of Chaitin’s assertion that measures mutual information: “ is simply the minimum attainable mutual information over any partition of into non-overlapping pieces all of size not greater than [Chaitin 1990b, 100).” But the mutual information between and for a “gas” scenario (random sequence), for instance, is at a maximum when . Perhaps Chaitin means simply that if measures organization it must indirectly measures mutual information as well, on the assumption that mutual information measures organization. [↑](#footnote-ref-11)
12. See especially Dembski 1998, Dembski and Ewart 2023. [↑](#footnote-ref-12)
13. The authors have, however, issued a critique of Chaitin’s attempt to model Darwinian evolution in terms of algorithmic information theory. See Ewart, Dembski, and Marks 2013b. [↑](#footnote-ref-13)
14. Here we present the original version of the flow chart in Dembski 1998. [↑](#footnote-ref-14)
15. For the design inference, we must take into account also what Dembski calls probabilistic resources, both “replicational” and “specificational” (see Dembski 1998, 175-184). In our example, the replicational resources are the number of chances of getting a royal flush (number of hands dealt) and the specificational resources are the number of ways of dealing a royal flush (four ways, one for each suit). [↑](#footnote-ref-15)
16. Fitelson, Stephens, and Sober (1999) and Ellsberry and Shallit (2011), for example, are representative of such criticisms, which sometimes appear motivated more by the desire to discredit Intelligent Design theory than to address flaws in the concept of specified complexity. Elsberry and Shallit, for instance, take Dembski to task for allegedly discounting the possibility that a given pattern might arise due to natural causes, citing Jocelyn Bell’s 1967 discovery of pulsars (Elsberry and Shallit 2011, 243). But Dembski’s explanatory filter specifically allows for such explanations (chemical crystals, for instance) and Dembski rightly notes that our being sometimes unable to rule out natural necessity as an explanation does not imply that no cases exist where we can easily rule it out (Dembski 2002, 14). Moreover, the concept of specified complexity is not necessarily linked to intelligent design in biology. [↑](#footnote-ref-16)
17. Not all algorithmically compressible sequences are patterned (the sequence of digits in the number , for instance, is algorithmically compressible but not patterned). [↑](#footnote-ref-17)
18. Dembski and Ewart (2023, 146ff) call this a “prespecification.” [↑](#footnote-ref-18)
19. A further difficulty in Dembski’s original account of the explanatory filter is his definition of complexity in terms of a .5 prior probability of an event occurring, given the probabilistic resources. That is, if, given the relevant probabilistic resources (number of repetitions and number of possible specifications) a specified event is unlikely then, according to Dembski, the event is by definition complex and we can reasonably conclude it did not occur by chance (Dembski 1998, 190-198). It would be desirable to define complexity independently of a specific threshold of unlikeliness. Moreover, prior probability tells us solely what the odds are beforehand that an event will occur by chance, not the odds, the event having occurred already, that it occurred by chance. For a design inference we must compare possible explanations and they all may be of less than .5 prior probability. Under what conditions may my fellow poker players reasonably conclude that I cheated when I deal myself, say, four of a kind twice in a row? Friends who know me well will dismiss the possibility of cheating as inconsistent with my moral character, even if the prior probability of the sequence of deals is well below .5. But perhaps a group of strangers would reasonably suspect cheating. In general, unlikely specified events occur frequently by chance without unduly surprising us or triggering a design inference. In later writings (Dembski 2005, Dembski and Ewart 2023), Dembski proposes an updated definition of specification in terms of short description length. We shall consider in 2.4 below whether this revision is successful. [↑](#footnote-ref-19)
20. In this section we consider the concept of algorithmic specified complexity (ASC) as first set forth in the papers by Dembski and coauthors Ewart and Marks (Ewart, Dembski, and Marks 2013a, 2014, 2015a, 2015b). Dembski and Ewart’s newly revised (2023) version of ASC we consider in section 2.4. [↑](#footnote-ref-20)
21. The authors themselves point out this resemblance in passing (Ewart et al. 2015a, 886). [↑](#footnote-ref-21)
22. The 205 bits for the compressed sequence is obtained by taking the 43 characters in the sentence times the Shannon information per character, assuming equal probability of letters of the alphabet (which is not true of English but permissible for illustrative purposes): . I have simplified this example by disregarding the program length and the information required to encode of the number of characters in the sentence. The actual figure Ewart et al. obtain for is . [↑](#footnote-ref-22)
23. In their analysis of the random scenario, where they claim to detect negative algorithmic specified complexity, the authors inexplicably fail to assign to the UTF-32 encoded value of 1,376 bits, instead assigning the value of or 204.46. Clearly, to compare the two strings we need to encode them both in UTF-32. There is, therefore, no mathematical detection of specified complexity in the case of the meaningful sentence compared to the random string. The two sequences are equally compressible and yield the same algorithmic specified complexity. [↑](#footnote-ref-23)
24. The general equivalence of optimally coded Shannon information and algorithmic information is clear in a case like this (on the equivalence see, for example, Adriaans 2008, 146–150). But the equivalence does not hold in every case. The number , for example, is algorithmically compressible, but since it contains no patterns it is already optimally coded in the sense of Shannon information. [↑](#footnote-ref-24)
25. In their more recent analysis of algorithmic specified complexity (Dembski and Ewart 2023), Dembski and Ewart do in fact discuss an example of this type, for which see section 2.4 below. [↑](#footnote-ref-25)
26. See Dalkilic et al. (2006) for a computer program able on such basis to distinguish authentic (meaningful) text from inauthentic text. [↑](#footnote-ref-26)
27. On this point Yockey 1992, 82: “[B]oth random sequences and highly organized sequences are *complex* because a long algorithm is needed to describe each one. Information theory shows that it is *fundamentally undecidable* whether a given sequence has been generated by a stochastic process or a highly organized process.” [↑](#footnote-ref-27)
28. Not all redundancies in DNA coding allow for algorithmic compression, however. For example, since the DNA code assigns sixty-one codons to twenty amino acids, most amino acids are assigned more than one codon. Assuming equiprobability of amino acids, this would result in an information loss of 1.61 bits, from 5.93 bits per codon to 4.32 bits per amino acid. Since amino acids are not equiprobable there is an even greater information loss. Moreover, when the substitutability of amino acids is taken into account—since some protein sites accept more than one amino acid—there is additional information loss. The protein iso-1-cytochrome c, for example, contains an average of only 2.11 bits per amino acid by Yockey’s calculation (1992, 131–172). The preceding biomolecular redundancies, however, do not allow for algorithmic compression because they are not based on algorithmically detectable patterns. [↑](#footnote-ref-28)
29. The authors are perplexed by this, for “one would ascribe to life a high complexity and a high degree of structure, and these criteria are certainly not fulfilled by a completely random string” (Schmitt and Herzel 1997, 375). The explanation of the phenomenon, the authors suggest, is that the expected redundancy in the DNA sequence is transferred to the redundancy of the translation mechanism, by which 20 amino acids are represented by 61 codons. But that is irrelevant since we should in fact expect a DNA sequence to exhibit high algorithmic randomness, which is equivalent to high algorithmic complexity. The fallacy, once again, is to assume that complexity is the opposite of randomness in information-theoretic terms. [↑](#footnote-ref-29)
30. Abel and Trevors (2005) correctly distinguish what they call “functional sequence complexity” from “random sequence complexity,” but they erroneously attempt to define the former in terms of a combination of complexity and compressibility, citing in this connection, however, solely the example of redundancy in written text—the “qu” sequence in English, for instance. But the randomness deficiency of an algorithmically complex sequence by virtue of some degree of associated redundancy merely defines a point on the continuum of order and randomness, not a third algorithmic-theoretic category alongside order and randomness. [↑](#footnote-ref-30)
31. Dembski seems to have first broached the idea in Dembski 2005. [↑](#footnote-ref-31)
32. Strictly speaking the specific words would have to be specified—but 240 bits is close enough. [↑](#footnote-ref-32)
33. See Yockey 2000, 119–120. This estimate assumes about 2,000 enzymes for metabolism and reproduction at about 374 bits per enzyme. [↑](#footnote-ref-33)
34. According to Yockey, no advocate of intelligent design, it is impossible for life to have originated by chance since biological information is not derivable from physics and chemistry (Yockey 2005, 2). This does not, on Yockey’s view, rule out neo-Darwinian evolution. [↑](#footnote-ref-34)
35. This is true in general, but particular examples can be cited, such as that the specific assignment of amino acids to DNA codons is determined by enzymes, that the cell can read a DNA sequence in different frames, that a DNA codon can specify a different amino acid in the mitochondria than in the nucleus of the cell, and so forth. [↑](#footnote-ref-35)