

# AN ALGORITHMIC INFORMATION THEORY CHALLENGE TO INTELLIGENT DESIGN

by Sean Devine

*Abstract.* William Dembski claims to have established a decision process to determine when highly unlikely events observed in the natural world are due to Intelligent Design. This article argues that, as no implementable randomness test is superior to a universal Martin-Löf test, this test should be used to replace Dembski's decision process. Furthermore, Dembski's decision process is flawed, as natural explanations are eliminated before chance. Dembski also introduces a fourth law of thermodynamics, his "law of conservation of information," to argue that information cannot increase by natural processes. However, this article, using algorithmic information theory, shows that this law is no more than the second law of thermodynamics. The article concludes that any discussions on the possibilities of design interventions in nature should be articulated in terms of the algorithmic information theory approach to randomness and its robust decision process.

*Keywords:* algorithmic entropy; algorithmic information theory; fourth law of thermodynamics; Intelligent Design; randomness test

---

William Dembski, in a number of works, including *The Design Inference* (1998), *No Free Lunch* (2002b), and "Specification: The Pattern that Signifies Intelligence" (2005), claims that there is a robust decision process that can determine when certain structures observed in the natural world are the product of Intelligent Design (ID) rather than natural processes. As defined by the Discovery Institute Web page (Discovery Institute 2012), the theory of ID "holds that certain features of the universe and of living things are best explained by an intelligent cause, not an undirected process such as natural selection. Through the study and analysis of a system's components, a design theorist is able to determine whether various natural structures are the product of chance, natural law, intelligent design, or some combination thereof."

Dembski also introduces his fourth law of thermodynamics that effectively states that *information*, using his definition, cannot increase by

Sean Devine is a research physicist who holds the position of Research Fellow at the School of Management, Victoria University of Wellington, Rutherford House, 23 Lambton Quay, Pipitea Campus, Wellington 6140, New Zealand; e-mail: sean.devine@vuw.ac.nz.

natural processes (2002b, section 3.10). He then argues that structures that are high in information cannot emerge by chance.

The essence of the first of these claims is that a robust decision process can be used to determine whether an observed event, that is, a structure, such as the flagellum that provides motility to certain bacteria (see Behe 1996), is an outcome of evolutionary processes, or is the product of nonnatural design. The Dembski decision process considers first whether such a structure or event can be explained by natural laws. If not, a randomness test is devised based on identifying and specifying the event  $E$ . When the probability of such a specified event occurring by chance is low, it is said to exhibit *Complex Specified Information* (CSI). According to Dembski, this event can be deemed to be due to ID as chance is eliminated. For example, Dembski (2002b, xiii) would see a random set of Scrabble pieces as complex but not specified, while a simple word “the” is specified without being complex. In contrast, a Shakespearean sonnet is both complex and specified and would be unlikely to occur by chance.

In mathematical terms, if  $P(E|H)$  is the probability of the specified event, given the chance hypothesis  $H$ , Dembski defines the *information* embodied in the outcome by  $I_D = -\log_2 P(E|H)$ . Dembski has defined his information measure so that the lower the probability of an observed outcome, the higher is the information and order embodied in the structures and, in Dembski’s terms, the higher the complexity. This measure is the converse of the usual mathematical definitions of information and here will be denoted by  $I_D$  and termed *D-information*. As is shown later, the concept of D-information has much in common with Kolmogorov’s *deficiency in randomness*, that is, just like the deficiency in randomness, outcomes with high D-information would exhibit low algorithmic information, low entropy, and low algorithmic complexity. Shallit and Elsberry (2004, 134–35) have noted the same point and have suggested that the term *anti-information* be used to distinguish the common understanding of information from what here is called D-information.

This article makes the following main points:

- As Shallit and Elsberry have suggested (2004, 134), Kolmogorov’s deficiency in randomness provides a far more satisfactory measure for D-information than that proposed by Dembski.
- As the Dembski approach does not adequately define a randomness test that can be implemented in practice (Elsberry and Shallit 2011), it should be replaced by the agreed mathematical measure of randomness known as a universal Martin-Löf randomness test. The universal randomness test achieves Dembski’s purpose and avoids all the confusion and argument around the Dembski approach.

- The clarity of the Martin-Löf approach shows that the Dembski decision process to identify ID is flawed, as the decision route eliminates natural explanations for surprise outcomes before it eliminates chance. The fundamental choice to be made, given the available information, is not whether chance provides a better explanation than design, but whether natural laws provide a better explanation than a design.
- Dembski's fourth law of thermodynamics, that is, his law of conservation of the information  $I_D$ , is no more than the second law of thermodynamics in disguise. It is equivalent to the unsurprising statement that entropy can only be conserved or increase in a closed system. Given the initial state of the universe, there is no evidence that the injection of D-information or its equivalent, the injection of low entropy from a nonnatural source, is required to produce any known structure.
- Dembski's claim that his law of conservation of information proves that high D-information structures cannot emerge by chance is irrelevant in an open system, such as the earth.

Elsberry and Shallit (2011) and Shallit and Elsberry (2004) provide a detailed critique of the inconsistencies of Dembski's idea of CSI and his so-called proof of the Law of Conservation of Information. While they and a number of other authors (Miller 2004, Musgrave 2006) show the bacterial flagellum can plausibly be explained by natural processes, as the ID supporters are likely to find other examples that they claim exhibits ID, the framework of the ID argument needs to be critiqued. Here, the primary concern is to show the whole mathematical approach used by Dembski is flawed. The mathematically robust Martin-Löf universal randomness test is used to replace Dembski's approach in order to determine whether natural laws can explain surprise events. This is particularly important, as influential thinkers, such as William Lane Craig, have been seduced by the apparent sophistication of the Dembski argument (Elsberry and Shallit 2011, 2). No implementable test of randomness can do better than a universal Martin-Löf test (Li and Vitányi 2008, 137). If order is recognized, the lack of randomness can be measured by this test. As a robust universal test of randomness (and therefore of order) already exists, the scientific community should only engage in discussions on the possibilities of design interventions in nature that are articulated in terms of this universal test.

#### ALGORITHMIC INFORMATION THEORY (AIT)

The key idea is that an ordered system can be more simply described than a disordered system. The AIT approach formalizes this by using a program or an algorithm as the means to describe or specify the system. The shorter

the program, the more ordered the system. For example, the sequence formed by drawing 100 balls from a lotto urn must specify each digit. On the other hand, the first 100 digits of  $\pi$  can be generated by a relatively short computer program.

The detailed mathematical treatment below is not for every reader. It deals with issues such as computer dependence and coding methodologies. However, the following are the key messages:

- (1) A system can be represented by a binary string in the system's state space. As an example, consider how one might represent the instantaneous configuration of  $N$  players on a sports field. If jumping is to be allowed, three position and three velocity dimensions are required to specify the coordinates for each player. With  $N$  players, the instantaneous configuration is a binary string in what would be called the six  $N$ -dimensional state space of the system. If the configuration is ordered, as would be the case if all the players were running in line, the description would be simpler than the situation where all players were placed at random and running at different speeds and directions.
- (2) Let  $s$  represent the string of binary characters that specifies the instantaneous configuration or structure of a system (such as the positions and velocities of the players as outlined in the previous paragraph). If a particular structure shows order, features, or pattern, as is the case for structures that might be considered to exhibit ID, a short algorithmic description can be found to generate the string  $s$ . Basically ordered structures have short algorithmic descriptions, disordered structures do not.
- (3) When appropriately coded, the length of the shortest algorithm corresponds to an entropy measure denoted by  $H(s)$ .
- (4) The measure of order in a particular structure is the difference in length between its short algorithmic description and the description of a disordered or random structure. This measure is called the deficiency in randomness and is measured in bits.

A simple example might be that of a magnetic material, such as the naturally occurring lodestone. The direction of the magnetic moment or spin associated with each iron atom can be specified by a 1 when the spin is vertically aligned and a 0 when the spin is pointing in the opposite direction. Above, what is called the Curie transition temperature, all the spins will be randomly aligned and there will be no net magnetism. In which case, the configuration at instant of time can be specified by a random sequence of 0's and 1's. However, where the spins become aligned below the Curie temperature, the resultant highly ordered configuration is

represented by a sequence of 1's and, as shown below, can be described by a short algorithm.

AIT provides a formal tool to identify the pattern or order of the structure represented by a string. The algorithmic complexity (or information content) of a structure represented by a binary string  $s$  is defined by the length of the shortest binary algorithm that is able to generate  $s$ . If this algorithm is much shorter than the string itself, one can conclude the structure represented by the string is ordered. However, as is outlined below, to be consistent, standard procedures are required to specify the structure as a string, to code the algorithms that describe the structure, and to minimize the computer dependence of the algorithm.

The basic concept of AIT was originally conceived by Solomonoff (1964). Kolmogorov (1965) and Chaitin (1966) formalized the approach and were able to show that the computer dependence of the algorithmic complexity can be mostly eliminated by defining the algorithmic complexity or information content of the string  $s$  as the length of the shortest algorithm that generates  $s$  on a reference universal Turing machine (UTM). A UTM is a simple general-purpose computer with expandable memory. Importantly, the universe is itself a UTM, and physical laws determine the computational path of the states of the universe. As a UTM can simulate any other Turing machine (Chaitin 1966, Li and Vitányi 2008), the reference UTM can in principle simulate the universe, or any other UTM, by taking into account the machine dependence of algorithms.

Consider the following two outcomes resulting from the toss of a coin 200 times where heads is denoted by a 1 and tails by a 0.

- (1) A random sequence represented by 200 characters of the form "110010 . . . 1100." This sequence can only be generated by a binary algorithm that specifies each character. If the notation  $|\dots|$  is used to denote the length of the binary string representing the characters or computational instructions between the vertical lines, the length of the program that does this is

$$|p| = |110010 \dots 1100| + |OUTPUT \text{ instruction}| + c.$$

As the length of this algorithm must include the length of the sequence, the length of the *OUTPUT* instruction, and a constant term reflecting the length of the basic instruction set of the computer implementing the algorithm, it must be somewhat greater than the sequence length.

- (2) The outcome is ordered consisting of 200 heads in a row, represented by the sequence "111 . . . 111."

The algorithm that generates this is *OUTPUT* 1,200 *times*. In this case, the algorithm only needs to specify the number 200, the character printed,

together with a loop instruction that repeats the output command, and again the constant  $c$  that includes the basic instruction set of the computer. That is, the length of the algorithm  $p'$  is:

$$|p'| = |200| + |1| + |OUTPUT\ instruction| + |loop\ instruction| + c.$$

This is somewhat greater than the 8 bits required to specify the integer 200. In what follows,  $p^*$  will be used to denote the shortest program that generates string  $s$ . As the algorithms  $p$  and  $p'$  above may not be the shortest possible,  $|p^*| \leq |p|$ . In general, the length  $|p^*|$  of the shortest algorithm that generates the sequence is known as the *algorithmic complexity*, the *Kolmogorov complexity*, or the *program-sized complexity* of the string. When appropriately coded, as is outlined below, this measure is also the algorithmic entropy of the configuration the string specifies. Any natural structure that shows order can be described by a short algorithm compared with a structure that shows no order and which can only be described by specifying each character.

However, there are two types of codes that can be used for the instructions and algorithm. The first is where end markers are required for each coded instruction to tell the computer when one instruction finishes and the next starts. This coding gives rise to the plain algorithmic complexity denoted by  $C(s)$ . Alternatively, when no code is a prefix of another, the codes can be read at any instant requiring no end makers. This requires the codes to be restricted to a set of instructions that are self-delimiting or come from a prefix-free set (Levin 1974, Chaitin 1975). The algorithmic complexity using this coding will be denoted by  $H(s)$  and, because it is an entropy measure, will be termed the algorithmic entropy.

The two complexity measures differ slightly by a term of the order of  $\log_2 C(s)$  (Li and Vitányi 2008, 203), that is,  $H(s) - C(s) \leq 2\log_2 C(s) + O(1)$ . Much of the discussion on testing for randomness can use either definition of complexity. While  $C(s)$  is more straightforward, in later discussions on information, and the so-called fourth law of thermodynamics,  $H(s)$  is more appropriate as it is an entropy measure that aligns with the traditional concept of entropy.

The formal definition of the algorithmic complexity first specifies that the computation using program  $p$  is implemented on a reference UTM  $U(p)$ . When there is no restriction on the computer instructions, the complexity measure is the plain algorithmic complexity  $C_U(s)$  given by:

$$C_U(s) = |p^*| = \text{minimum}|p|\text{such that } U(p) = s.$$

Similarly,  $H(s)$  is given by virtually the same equation, but there is the additional requirement that no instruction in  $p$  can be a prefix of any other.

As different UTMs can simulate each other, the algorithmic complexity measure on a particular machine can be related to another by a constant term of the order of 1, denoted by  $O(1)$ . This allows the machine-independent definition to be given by:

$$C(s) \leq C_U(s) + O(1)$$

$$H(s) \leq H_U(s) + O(1)$$

When a simple UTM is used, the  $O(1)$  term will be small, as most instructions can be embedded in the program rather than in the description of the computer.

In many physical situations, only the differences between the lengths of algorithms are the important measure. In this case, the  $O(1)$  term cancels out and common instructions, such as the *OUTPUT* instruction, or those specifying natural laws, can be taken as given and ignored. A further point is when the computation starts with an input string  $t$ , the algorithmic complexity is denoted by  $C(s|t)$  or  $H(s|t)$ .

Ignoring common instructions and machine dependence, the algorithmic complexity of the random string above becomes:

$$H(110011 \dots 110) \approx C(110011 \dots 110) \text{ is given by.}$$

$$|p^*| \approx |110011 \dots 110|.$$

On the other hand, the ordered string of 200 heads is represented by

$$H(111 \dots 111) \approx C(111 \dots 111) \text{ is given by } |p^{*/} \approx |200| + |1|$$

$$+ |\text{loop instruction}|.$$

$C(110011 \dots 110)$  is more than 200 bits, whereas  $C(111 \dots 111)$  is much shorter as it only requires about 8 bits to specify the integer 200. There are only a few more bits to account for the loop instruction, so on. The specification of the ordered string is close to  $192 = 200 - 8$  bits shorter than the random string. Kolmogorov introduced the term *deficiency in randomness* to quantify the amount of compression. While the algorithmic complexity is not computable, where order is recognized, the description can at least be partially compressed. If more hidden structure is found, the description can be compressed further.

*Nomenclature—What is Information?* The Nobel laureate Manfred Eigen recognized that the nucleic acids code information. This leads William Dembski to argue that information is key to unraveling the central problems of biology (2002a, b) and to claim an injection of information is necessary for living systems to reach certain levels of biological complexity.

However, there is no reason to believe that Dembski's  $I_D$  or D-information corresponds to what Eigen meant. While there is more than one way to define information, it is important to be consistent and to understand how different definitions are related. Dembski justifies his D-information concept by comparison with Shannon's information theory for a message transmitted by a source, through a communication channel to a receiver. The received message is the event  $E$ . The amount of information transmitted, according to Dembski, is given by  $I_D = -\log_2 P(E|H)$ , assuming the chance hypothesis  $H$ . This definition assigns higher information content to more highly ordered structures that might exhibit CSI.

On the other hand, Shannon's Information Theory defines *information* as the number of bits required to identify a particular message in a set of messages. In this approach, the length of the optimum code for each message is virtually  $-\log_2 P(E|H)$ . While the expected outcome of D-information for a set of outcomes is the same as for Shannon information, Shannon warns that "the concept of information applies not to the individual messages but rather to the situation as a whole" (Shannon and Weaver 1949, 100).

In the Shannon Information Theory approach, as the number of messages increases, the information increases. While the AIT approach must actually define the message or string, its definition of information aligns with that of Shannon. As a consequence, AIT identifies increasing information with increasing disorder, greater randomness, and increasing entropy. Because ordered structures can be defined by short algorithms, in contrast to Dembski's definition, the information content and the algorithmic complexity of these is low. In the algorithmic case, the information is embodied in the number of computational bits needed to define the system. This is lower for ordered systems.

Most structures in the living biosystem are highly ordered and far from even a local equilibrium. Each of them can be described by an algorithm that has fewer bits relative to the description of an equilibrium configuration. A tree is such a case, as it can be specified by the growth instructions embodied in the DNA of the seed cell and the environmental conditions affecting the growth algorithm. The amazingly real looking fractal models of natural structures for use in animated movies or creating artificial ecosystems, demonstrate the ability of simple algorithms to capture the richness of what intuitively would seem to be immensely complex natural systems.

Dembski's use of the word *complexity* is also confusing (see also comments in Elsberry and Shallit 2011, Shallit and Elsberry 2004). At times he identifies increasing complexity with increasing randomness, such as when he compares a Caesar cipher with a cipher generated by a one-time pad (Dembski 2002b, 78). At other times, he identifies increasing complexity with increasing order (Dembski 2002b, 156–83).



The AIT approach does not fall into the ambiguity trap of Dembski's definitions. Furthermore, as is discussed below, the deficiency in randomness provides an information measure with exactly the properties required by Dembski. The deficiency measure leads to an understanding of information that makes it clear that there is no need for any fourth law of thermodynamics, as there is no need for an injection of information to generate currently observed living systems.

#### DEFICIENCY IN RANDOMNESS AS A MEASURE OF ORDER

A disordered system requires a long description as each component needs to be separately specified. On the other hand, if the system is ordered, a short algorithmic description is possible. The difference in bits between a binary specification of a random string and a string of the same length that shows order is called the *deficiency in randomness*. It is a measure of how nonrandom or how ordered the particular configuration is relative to a disordered one. For example, if the system is the players on the sports field, the difference in length between the description of the configuration of players where their positions and velocities are randomly distributed and one where the players are lined up is the deficiency in randomness. The following formalizes the approach.

In AIT, the word *typical* is used to describe a string that is deemed random in a set of strings, that is, it has no distinguishing features. Similarly, a typical string in a real-world system would be deemed to belong to a state in the set of equilibrium states. As mentioned above, the deficiency in randomness of string  $s$  is a measure of how untypical or how ordered a string is and, as discussed later, forms the basis of a universal Martin-Löf randomness test.

Much of the ID discussion is about identifying nonrandom strings in a set of equally likely outcomes. The probability distribution over the members of this set is called the *uniform distribution*. The algorithmic complexity of a typical string  $s$  in the  $2^n$  members of the set of all strings with length  $n$ , is the length of the string itself, that is,  $|s| = n$ . If all strings are equally likely this is the same value as the Shannon entropy of the set. That is, the algorithmic complexity of a typical or random member of the set = the Shannon entropy of the set.

The deficiency in randomness can be defined using either plain algorithmic complexity  $C(s)$ , or  $H(s)$  for self-delimiting coding. As  $H(s) - C(s) \leq 2\log_2 C(s)$ , the difference between these is usually unimportant, although the discussion is a little simpler using plain complexity. However,  $H(s)$  has the advantage that it can be identified with the thermodynamic entropy (allowing for units). When  $H(s)$  is used, the extra length given by  $O(\log_2 C(s))$  term may need to be tracked.

The deficiency in randomness for string  $s$  where  $|s| = n$  for simple coding is defined as

$$d(s|n) = n - C(s|n).$$

Here,  $C(s|n)$  is the plain algorithmic complexity given the value of  $|s| = n$ . The deficiency is close to 0 for a typical or random string, and approaches  $n$  for a highly ordered string. Similarly, for self-delimiting coding and  $P(s) = 1/n$ , the deficiency is defined as  $\delta(s|P(s))$ ,

$$\delta(s|P(s)) = n - H(s|n).$$

For a real-world system, the self-delimiting definition measures the distance string  $s$  is from an equilibrium string. In the case of a general distribution  $P(s)$ ,

$$\delta(s|P(s)) = -\log_2 P(s) - H(s|P(s)).$$

In what follows, the major contributions to the deficiency of randomness will be outlined to illustrate the idea. For the sake of a more straightforward argument, any small  $O(1)$  remaining terms and the  $O(\log_2)$  extra term arising from self-delimiting coding will be ignored.

Consider the highly ordered outcome of obtaining 200 heads in a row from the toss of a coin 200 times. This can be specified by an algorithm slightly greater than 8 bits. On the other hand, as all outcomes are equally likely, a typical or random outcome requires at least  $n = 200$  bits to be specified. The deficiency in randomness is the difference between these and is close to 192 ( $= 200 - 8$ ) bits. The large deficiency in randomness indicates the outcome is a surprise, as it is extremely unlikely to be due to chance. Such a surprise outcome has low algorithmic complexity or low algorithmic entropy, and represents a high degree of order.

In the section headed “The Universal Randomness Test for Design,” it is shown that the amount the description of an outcome can be compressed is the basis of the mathematically robust universal Martin-Löf test of randomness. As this test is the most reliable tool to identify the level of randomness in a given string, it should replace the Dembski design filter. The masterstroke of the Martin-Löf approach is that a universal test cannot be bettered, and any workable test can always be expressed as a universal test. Furthermore, in the section on “Entropy Information and a Fourth Law of Thermodynamics,” the deficiency in randomness using self-delimiting coding, provides an alternative measure of D-information and shows the relationship between D-information and (algorithmic) entropy.

*Limitations of Deficiency in Randomness.* Although no computable process can unequivocally determine the extent of the pattern or order in an observed outcome, this is not so critical for the ID situation. As the order or pattern is recognized by observation, a provisional algorithm can be found to capture this order. Even if it is not the shortest algorithm possible, the recognized structure may provide a sufficiently good estimate of the degree of order to address the critical ID questions and would satisfy all Dembski's requirements.

Nevertheless, if there is a need to find a shorter algorithm, the resource-bound complexity can be used as an approximation to the algorithmic complexity, providing a lower bound to the deficiency in randomness. This approximation is just the shortest algorithm that generates the string in no more than  $t$  steps.

#### DEMBSKI'S DECISION PROCESS TO IDENTIFY ID

Dembski (1998) claims to have developed a robust decision process that determines whether chance or, alternatively, a nonnatural intervention explains certain observed events. This decision process focuses on differentiating low-probability events that occur by chance, from similarly low-probability events that can be shown to exhibit what Dembski calls *CSI*. As Dembski explains, "A long sequence of randomly strewn Scrabble pieces is complex without being specified. A short sequence specifying the word "the" is specified without being complex. A sequence corresponding to a Shakespearean sonnet is both complex and specified" (Dembski 2002b, xiii). In essence the process is trying to distinguish a highly ordered surprise outcome from a random one.

Once the outcome has been shown to be specified, the results can be fed into a decision process that Dembski calls the *design filter*. If the design filter shows that natural laws cannot explain the specified outcome, and it has low probability, according to Dembski it cannot be explained by chance and therefore must be the result of ID. The logic of the argument outlined in the left-hand column of Table 1 is as follows:

- (1) Can an observed outcome  $E$ , for example, a highly complex biological structure, be explained by the regularity of natural laws?
- (2) If not, is the outcome due to chance? That is,
  - Is the probability of the outcome independent of what Dembski calls background information,  $K$ ? This independence is to ensure that the background information does not change the probability of the outcome. This requires that  $P(E|H, K) = P(E|H)$ , where  $H$  is the chance hypothesis.
  - Can the outcome be specified? The process of specification has evolved through Dembski's writings. Essentially it involves,

**Table 1.** Comparison of Dembski and AIT Decision Processes

Dembski decision process	AIT decision process
<b>Does a natural explanation exist?</b>	<b>Ignore: this should be last step</b>
Is the event $E$ chance?	Is the surprise event $E$ (represented by $s_D$ ) due to chance?
Is $P(E H, K) = P(E H)$ ?	
Does pattern $D$ delimit $E$ , and can $D$ be specified, and is the probability low even considering all the probabilistic resources?	If in a Martin-Löf test, $d(s_D) - 1$ or $\delta(s_D) - 1 \geq m$ for large $m$ , then chance probability $P(s) \leq 1/2^m$ is low.
That is, is $P(\text{overall}) < 1/2$ ?	
If yes, $D$ shows CSI	
<b>Therefore not chance but ID</b>	<b>Not chance as structure is ordered</b>
	Is a natural explanation in principle possible?
	That is, can order emerge through the ejection of disorder, perhaps using stored energy?
	(see text)

given the background information: Can a general complexity measure  $\varphi$  be found to assign  $E$  to a rejection region of possible outcomes? That is, a region where chance can be eliminated (Dembski 1998, 2002b).

- Is the probability of  $E$  still low, taking into account all the probabilistic resources (i.e., the myriad of different opportunities for  $E$  to occur and to be specified)?
  - If the resultant probability that includes all the chance ways of producing  $E$  is  $< 1/2$ , then according to Dembski, chance as an explanation is eliminated.
- (3) If such an unlikely outcome is observed, it embodies information that is both complex and specified and, according to the Dembski procedure, one must conclude that the outcome exhibits design.

As Dembski's decision process eliminates natural laws as the first decision step, it "privileges design as an explanation" (Elsberry and Shallit 2011). Hence, it will assign design to structures that are poorly understood. The choice should be between natural laws and design not chance and design.

Dembski's arguments have evolved (1998, 2002b, 2005), presumably because of weaknesses in earlier approaches. As no workable example of the process is given, there are difficulties in applying it—only possibilities are suggested. In the Design Inference Dembski (1998) defines an abstract function  $\varphi$  to specify the event that might exhibit design, together with an argument based on the likelihood of the event occurring by chance over many observations. *No Free Lunch* focuses rather on Fisher's rejection of region  $T$  to eliminate a chance event. Dembski recognizes (1998, 167;

2002b, 61), the resource-bound algorithmic complexity could be used to specify the outcome, but for Dembski it is just one of several ill-defined possibilities. Later, Dembski (2005) gets remarkably close to the Martin-Löf approach by using a specification process based on how much an algorithmic description can be compressed. However, he then vaguely defines a *descriptonal complexity* measure for a general outcome. But rather than using the Martin-Löf approach, he feeds this into Fisher's statistical significance test.

This further suggests, as is discussed below, that if a workable version of the Dembski approach is to be found, it will end up using the deficiency in randomness. In which case the measure should be fed into the Martin-Löf randomness approach, rather than Dembski's design template.

### THE UNIVERSAL RANDOMNESS TEST FOR DESIGN

The argument behind the universal randomness test can be illustrated by a simple example using the fact that, in general, there are only  $2^{k+1} - 1$  strings of length  $k$  or less. This provides an upper limit on the number of algorithmic descriptions shorter than  $k$ . Given a string of length  $n$  where say  $n = 8$ , how many strings can be generated by a program that is more than 3 bits shorter than the length 8 that is compressed by 4 bits or more? These are the strings of length 4, 3, 2, or 1 bits. No more than  $2^{8-3} = 32$  strings can be compressed this much. Of the 256 strings of length 8, only 32 can be compressed by 4 bits or more. One example is the string 11111111 as it can be simply described. While a random 8-bit string cannot be compressed, a string that can be algorithmically compressed by more than 3 bits is relatively ordered. But, as there are fewer than 32 such strings, the probability of getting such an outcome by chance is  $32/256$ . One can say that these short strings can be rejected as random at level 3. In general, a string is rejected as random at level  $m$  when the string can be compressed by  $m + 1$  or more bits. Given that there are  $2^n$  strings of length  $n$  fewer than  $2^{n-m}$  can be compressed by  $m + 1$  or more. These strings can be rejected as random at level  $m$  and the probability that it is possible to find such a compressed description is at most  $2^{n-m}/2^n = 1/2^m$ . As this argument shows there are limits to the number of ordered strings available, therefore the approach can become a test for randomness.

While there are many tests for randomness that satisfy the Martin-Löf criteria discussed below, the test function used here and illustrated in the previous paragraph is based on the deficiency in randomness. Furthermore, it is a universal Martin-Löf test. Basically, this test identifies the lack of randomness (and hence the level of order) in a string  $s$ , by how much the shortest algorithm that generates string  $s$  compresses it. Using the deficiency in randomness to quantify the amount a string can be compressed,  $d(s|n) \geq m + 1$ . It follows that  $d(s|n) - 1 \geq m$  bits. This allows a test

function that satisfies the Martin-Löf randomness test requirements to be defined as  $\text{Test}(s) = d(s|n) - 1$  bits.

The general Martin-Löf randomness test places two requirements on a test function  $\text{Test}(s)$  (Li and Vitányi 2008, 135), that is, that

- (1)  $\text{Test}(s) \geq m$ , and
- (2) The cumulative probability for all strings compressed more than  $m$ , must be no more than  $1/2^m$  for all  $n$ . Or equivalently, the number of such string is no more than  $2^{n-m}$ .

If these criteria are satisfied, the string  $s$  can be rejected as random (and therefore can be considered ordered) at level  $m$ . The first requirement is that the  $\text{Test}(s)$  identifies the greatest value of  $m$  that defines a set of ordered strings containing  $s$ . The second requirement ensures that the fraction or number of these strings the test selects is consistent. While there are many valid tests, the deficiency in randomness is the most suitable for addressing the ID question.

As an example, the outcome  $s = 200$  ones, generated by tossing 200 heads in a row, can be compressed to a program slightly more than 8 bits long, ignoring overheads (and if self-delimiting coding is used, the  $O(\log_2)$  term). That is, as a typical or random string would need 200 bits to be specified, the ordered string can be compressed by at least 192 bits giving  $\text{Test}(s) \geq 191$  with the test function  $d(s|n) - 1$ . The probability of a string being compressed to 8 bits or less is no greater than  $1/2^m = 1/2^{191}$ . It is therefore a valid test and can be rejected as random for any  $m \geq 191$ . Furthermore, because  $m$  is so large, the result 200 heads in a row is extremely unlikely to be observed by chance.

There is no clear distinction between random and nonrandom strings. Rather, the choice of the value of  $m$  at which randomness can be rejected defines the boundary between random and nonrandom (Martin-Löf 1966). The set of strings categorized by  $m$ , nests the less random strings with higher values of  $m$ . At  $m = 0$ , all strings are included below the cutoff and therefore none can be rejected as random. While at  $m = 1$ , no more than half the strings can be rejected as being random and so on.

The use of the algorithmic entropy  $H(s)$  provides for a more intuitive test process. This formulation, known as the sum  $P$ -test (Gács 2010, Li and Vitányi 2008, 278), is also a Martin-Löf test and is universal (Li and Vitányi 2008, 278–79). In essence, this procedure identifies an ordered string by a betting argument. Player A places a \$1 bet claiming that an awaited outcome, based on tossing a coin 200 times, will not be a random sequence as the coin toss is rigged. To test this, A requires that the payoff for the bet is  $2^\delta$ , where  $\delta$  is the difference between 200, the length of a random outcome, and the shortest algorithmic description of the string that eventuates. As is shown below, because the bet is structured so that the

expected return is no more than \$1, it is a fair bet. This means that a high return shows the outcome is a surprise and could not have happened by chance. The greater the level of surprise for such an outcome, the greater the payoff.

The proof that this will work is based on the requirement for self-delimiting coding, namely, the Kraft inequality, which requires that  $\sum (1/2^{H(s)}) \leq 1$ .

The bet takes a hypothetical payoff function, based on the self-delimiting deficiency in randomness, to be  $2^{\delta(s|P(s))}$ . This is a fair bet as the expected return for a \$1 bet is  $\sum P(s)/(2^{H(s)} P(s))$ , which is less than \$1 because of the Kraft inequality. In general, if a \$1 bet is placed assuming the uniform distribution, the payoff is  $\$2^{(n-H(s))}$ . The typical payoff in tossing a coin 200 times is expected to be  $\leq \$1$  while for the surprise outcome 200 heads in a row, the payoff is enormous at  $\$2^{192} = 10^{64}$ . Even if the outcome is a string that is compressed as little as 10 bits, the payoff is \$1,024. In general, if the outcome  $s$  eventuates, the largest value of  $m$  that satisfies  $\delta(s|P(s)) \geq m + 1$ , determines the payoff to be at least  $\$2^{m+1}$ . For random strings  $m \approx 0$ . Strings can be ranked by the payoff, the greater the payoff, the greater the value of  $m$ , and the more ordered is the string.

Particular randomness tests are said to be universal in that any computable randomness test (including any workable version of Dembski's test) can always be expressed as a universal test (Li and Vitányi 2008, 134). Furthermore, no computable randomness test, either known, or yet to be discovered, can do better than a universal test. As the deficiency in randomness, either in its simple form (Li and Vitányi 2008, 139), or in its self-delimiting form is a universal test (Li and Vitányi 2008, 279), the measure can be used to provide a robust decision process to identify nonrandom structures. Either test should replace Dembski's design filter as they both avoid the confusion of the Dembski test and the need for the *specification* concept or a process to eliminate chance. The deficiency in randomness fulfils all Dembski's specification requirements and provides an upper limit on the probability of a particular outcome in the set of possible outcomes. If a string can be rejected as random at a sufficiently large value of  $m$ , the probability of it occurring by chance will be less than  $1/10^m$ . Any string that is deemed not random for large  $m$ , exhibits what Dembski would call CSI as it has a much simpler description, and the probability that it would occur by chance is low. However, whether this implies ID or not, depends on whether a natural explanation is realistic.

#### COMPARISON BETWEEN THE DEMBSKI DESIGN TEMPLATE AND THE MARTIN-LÖF TEST

The left-hand column of Table 1 shows how the Dembski test aligns with the universal test shown in the right-hand column. Let  $s_D$  be the

uncompressed string representing the outcome  $E$ . Dembski denotes this by  $D$ . The question is whether this string would be a surprise outcome in the set  $A$ .

As was outlined in the section on Dembski's decision process, once a natural explanation is eliminated, the Dembski process attempts to establish whether the string  $D$  representing the outcome  $E$  can be specified and whether chance is eliminated. Chance must be eliminated by taking into account all the *probabilistic resources*, that is, all the opportunities for the outcome to occur and be specified. If the probability of the event occurring by chance is still extremely unlikely, the string is said to show CSI, in which case, Dembski argues CSI indicates ID.

On the other hand, the right-hand column of Table 1 recognizes that the decision in the end is between a design intervention and natural laws, not chance and design. In summary, the right-hand column in Table 1 involves the following steps:

- (1) Can chance explain this event; that is, is it random relative to a set of alternative outcomes using a universal randomness test? This involves determining how much the algorithmic description can be compressed. If the value of  $d(s_D|P(s_D)) - 1$ , or  $\delta(s_D|P(s_D)) - 1 \geq m$ , for large  $m$ , the string can be deemed to be nonrandom at that level of  $m$  and is unlikely to have occurred by chance.
- (2) If it is not a chance event, because the test indicates the structure is ordered or nonrandom, is there a possible natural explanation for the emergence of order? At this point it becomes apparent that most observed natural structures will show a high degree of order. The reason is that, as the universe is ordered at one time, at a later time most of the order remains but in a different form. Most natural outcomes will, therefore, appear nonrandom on a Martin-Löf test, but still emerge through natural processes. The critical question, which should be asked at this stage and no earlier, is whether a natural explanation is possible.

The only certain way to rule out a natural explanation is to demonstrate that the system cannot repackage existing structures and eject sufficient disorder to create more ordered structures. It is not a question of showing a surprise outcome occurs in nature, as surprise outcomes occur everywhere, but a question of showing that the surprise outcome is inconsistent with the second law of thermodynamics or the conservation of energy. The steps to do this are as follows.

Can the system eject sufficient disorder to leave an ordered structure behind? And/or does it have the capacity to access low entropy external resources or sources of concentrated energy, such as light, or chemical species that can generate the observed order and eject



disorder? In effect, the question is whether there is a sufficient flow through of order, and the right sort of energy that can be repackaged to create new forms of order—noting that overall the entropy of the system and its environment will increase. Even in the unlikely event, a low entropy source has not been identified, a natural explanation might still become obvious with a deeper understanding of the fundamental physics behind quantum theory and gravity, or better understanding of emergent properties.

- (3) If, and only if, an observed outcome has no natural explanation should a nonnatural explanation be considered.

The Dembski design filter fails as natural causes must not be eliminated before chance is ruled out. The choice is not between chance and design, but between nonnatural design and an explanation based on natural laws, recognizing, of course, that the laws themselves may involve natural selection processes acting on variations in structure.

An example of a specified, highly ordered structure that is unlikely to appear by chance in the life of the universe is magnetized lodestone (magnetite). Indeed, the probability of magnetizing a mole of lodestone by chance is equivalent to tossing something like 1,023 heads in a row. Nevertheless, because disorder as heat can be passed to the environment, at a temperature below the Curie temperature all the magnetic spins associated with each iron atom can align by natural processes. Despite the low probability of this outcome happening by chance, if one did not know the mechanisms behind the ordering, natural processes could still not be ruled out. Natural laws can and do create such structures by the ejection of disorder or high entropy waste. It might be argued that the appearance of magnetized lodestone could not be due to design because a natural explanation exists. Even if a natural explanation were not known, the entropy flow through indicates that such ordering is possible and may even in some circumstances become likely, as the critical driver of order is the ejection of disorder. While a particular ordered structure might be improbable in an interacting mix of structures at a local equilibrium, once entropy or disorder can be ejected, the structure becomes likely.

When a natural explanation is not known, the decision process should follow the outline of the right-hand side of Table 1.

*Difficulties with the Dembski Examples.* However, there are further serious conceptual problems with Dembski's approach. Van Till (2002), in reviewing Dembski's book *No Free Lunch* (2002b), points out that at times what Dembski calls the chance hypothesis *H* includes (1) chance as a random process; (2) necessity; or (3) the joint action of chance and necessity. As Dembski categorizes all three as stochastic processes (Dembski 2002b, 150), he is in effect claiming that his probability approach can,

and does, take into account all natural causes. Van Till (2002) points out that the comprehensiveness and inclusiveness of these terms must be understood in order to see the extremity of Dembski's numerous claims. Whatever Dembski might claim in theory, in practice he fails to take into account any cause except chance. For example, when Dembski attempts to establish whether the flagellum that provides motility to *Escherichia coli* (Behe 1996) exhibit CSI, he eliminates all natural causes except chance. Dembski describes the structure as a "discrete combinatorial object," and only tests the hypothesis that it is formed by the random alignment of the building blocks of proteins. As his  $P(E|H)$  does not take into account the most likely causal paths, the claim that such a structure is extremely unlikely to appear by natural processes is unsupported (see Elsberry and Shallit 2011, Shallit and Elsberry 2004, Miller 2004, Musgrave 2006). The evidence is that the observed structure in the Behe illustration can be plausibly explained by natural processes (Jones 2008).

Van Till (2013) highlights another critical flaw in Dembski's decision process. Dembski implies that biological evolution is about actualizing (or forming) a particular biological structure. On the other hand, a biologist's concern is how an evolutionary process might generate an adaptive function. As Van Till argues, the motility function of *Escherichia coli*, that is purported to be due to ID, is the critical feature, not its structure. Indeed, no biologist is interested in complicated structures that have no function, no matter how they might be formed. Consequently, the biological question is: "Can the motility function emerge through relatively small changes in the genetic code embodied in the ancestors of the bacterium?" In which case, if  $M$  denotes the motility function, given the  $N$  possible paths that might produce such a function, the probability that  $M$  will emerge becomes  $P(M|N)$ . This probability is likely to be many, many orders of magnitude greater than Dembski's  $P(E|H)$ , which is only the probability that the bacterium flagellum, as a structure, can be randomly assembled from its constituents parts.

Dembski similarly creates problems with specification. Both Dembski's mathematical definition of specification, and his nonbiological illustrations discussed below, imply that specification is about defining the observed structure with its pattern. Indeed, one would expect the specification of the structure to refer to something equivalent to its blueprint, its configuration, or some description of it. Yet, as Van Till (2002) points out, Dembski ignores his own mathematical development and states unequivocally that specification is about function (Dembski 2002b, 148). Similarly, Elsberry and Shallit (2011) note that for Dembski, function becomes a stand-in for specification. While the structure might imply certain functions, it is confusing to argue that one should specify the function rather than the structure. In other words, there is a strong argument that sees complexity as primarily related to function, and specification as primarily related to

structure rather than, as Dembski implies, the opposite. One suspects he does this because it is easier to provide a hand waving description of a function, than it is to adequately specify a structure.

Despite the rhetoric, the flagellum example discussed above, with all its flaws, is the only realistic illustration of what is purported to be ID (Elsberry and Shallit 2011). While Dembski uses two nonbiological examples to illustrate his process for recognizing design, because of the vague arguments, it is unclear how to operationalize his decision process, as the following shows.

The first nonbiological Dembski example involves a New Jersey official, Nicholas Caputo, who for a number of years was responsible for fairly ordering the political parties on a ballot paper. However, as the Democrats (as opposed to Republicans) appeared first place in 40 of the 41 ballot papers, it appears Caputo was cheating. In order to verify this, Dembski determines the probability that the Democrats should come first on the ballot papers at least 40 times. He then argues that the outcome of at least 40 Democrats in first place could not have occurred by chance, as it falls within a defined rejection region using Fisher's approach to hypothesis testing. Therefore, Caputo cheated. However, the real issue is: On what basis could a similar outcome, observed in the biological area, be deemed to be the result of an external intervention rather than natural processes? The example is of little help for this situation.

The second illustration is a fictional example from the movie *Contact*. In the movie, an extraterrestrial radio signal was found to list the primes up to 89 in ascending order. Dembski seeks to determine whether such an observation implies an intelligent source for the signal. Dembski states the probability of this happening by chance is  $1$  in  $10^{300}$ . He then claims that, as the maximum number of computations in the universe is  $10^{150}$ , such an outcome could not have happened by chance in the life of the universe. The trouble is that any other signal of the same length, even a random one, has the same probability and similarly could not have occurred. This is bizarre. On the other hand, the Martin-Löf approach does not fall into the Dembski trap, because it clearly distinguishes random outcomes from surprise outcomes.

In the end, as these nonbiological examples assume every outcome is equally probable, they are too simple to be helpful. On the other hand, for the flagellum case, the results depend critically on the assumed distribution and the precursors. Why go down the Dembski track when the Martin-Löf test avoids the ambiguities in the Dembski process? The decision question then can be articulated as: "Given the precursors, is this ordered structure improbable in the set of possible outcomes?" Also, the Martin-Löf process is mathematically robust and considerably simpler than the Dembski approach, while satisfying all his requirements. The integer  $m$  derived from the Martin-Löf test identifies the boundary between an ordered region

of outcomes and a disordered region, avoiding the need to use Fisher's statistical approach to reject the chance hypothesis. Critically, if the set of possible outcomes is appropriate, and the observed event corresponds to a high value of  $m$ , a reliable probability upper bound comes for free, that is,  $P(s_D) \leq 1/2^m$  for the rejection region  $m$ .

#### ENTROPY, INFORMATION, AND A FOURTH LAW OF THERMODYNAMICS

Dembski (2002a; 2002b, 166, section 3.10) introduces a law of conservation of information as a fourth law of thermodynamics. This implies that outcomes that exhibit CSI cannot be generated by natural causes. Effectively, the law is used to argue that as  $I_D$  is conserved, or can only decrease, high  $I_D$  outcomes that indicate design, cannot occur by chance. The Dembski argument for this law involves a somewhat convoluted discussion of the problems with Maxwell's demon (Dembski 2002b). Dembski seems unconvinced that demon paradox disappears once one recognizes that the demon is constrained by natural laws, as was pointed out by Landauer (1961), Bennett (1982, 1987), and Leff and Rex (1990). There also seems to be a circular argument. The appearance of CSI is used to demonstrate design and, therefore, the fourth law of thermodynamics. But surely one cannot then use this law to justify design?

Dembski's information concept also has difficulties (Shallit and Elsberry 2004, Elsberry and Shallit 2011). It is unable to adequately distinguish the level of surprise in an outcome of 200 heads in a row and an outcome of 180 heads mixed with 20 tails. Both are equally likely with  $I_D = 200$ . While both might exhibit CSI, the first outcome would be considered a far greater surprise than the second. Once it becomes clear that the deficiency in randomness is a better measure of D-information than the one used by Dembski, the meaning of the so-called fourth law emerges. Instead of the quantifying the amount of D-information in string  $s_D$  by  $I_D = -\log_2 P(E|H)$ , a modified measure  $\hat{I}_D$  is taken to be  $\hat{I}_D = |s_{\text{typical}}| - H(s_D)$ . The D-information contained in  $s_D$  becomes the number of bits separating the ordered string from a random or typical one. This makes sense as more highly ordered strings embody more D-information.

The relationship between D-information and entropy arises because  $H(s)$  is an entropy measure corresponding to the Shannon entropy for a set of outcomes (Bennett 1982). However, the relationship with the thermodynamic entropy can be made specific. Consider the situation where  $\Delta H$  bits shift a system from a macrostate, where each microstate has the same  $H(s)$ , to another set of states each with the algorithmic entropy  $H(s) + \Delta H$ . The equivalent heat flow into the system that has the same effect is  $k_B T \Delta H \ln 2$ , corresponding to a thermodynamic entropy flow of  $k_B \Delta H \ln 2$ .

The relationship between the D-information and the algorithmic entropy now makes it clear that the fourth law is just a rephrasing of the second law of thermodynamics. Rearranging the definition of  $\hat{I}_D$ , as  $H(s_D) = |s_{\text{typical}}| - \hat{I}_D$ , and as  $H(s_D)$  is an entropy; the second law of thermodynamics requires that  $H(s_D)$  can never decrease in a closed system. This requires that  $\hat{I}_D$  can never increase. As  $\hat{I}_D$  has the requisite properties of D-information, Dembski's claim that D-information can never increase by natural processes is equivalent to the second law of thermodynamics. What the second law of thermodynamics implies is that more order cannot arise from less order. If D-information is defined consistently, there is no need for a law of conservation of information. Current understandings of the second law explain all that so far needs to be explained. The next section shows how AIT allows one to track entropy and information at the scale of the universe.

*Order in the Universe.* Dembski (2005) takes Seth Lloyd's (2002) value of  $10^{150}$  (2002) as the maximum number operations of the universe. Dembski uses this as an upper value of the probabilistic resources, that is, the number of chances the universe has to produce a rare event. However, the argument is irrelevant, as the states in the universe are not independent, but highly correlated. This mistake seems to arise because Dembski has eliminated natural processes, therefore he must rely on chance. But once order appears in the universe, subsequent states will inevitably show some order.

Shortly after the Big Bang, the universe was in a highly ordered configuration. If the physical laws were completely known, the string representing this initial configuration would be highly compressed algorithmically in terms of these laws. However, because the algorithm that specifies a particular configuration of the universe at a later time must halt, the length of this algorithm, representing the algorithmic entropy, must contain information about the number of steps  $t$  to reach that halt configuration. Over time the number of steps increases and the algorithmic description grows as the steps to the halt state increases. This leads to an increase of the algorithmic entropy. Ultimately, at equilibrium, the shortest algorithmic description will be the one that specifies a random string in the set of equilibrium states. If physical laws are taken as given, this equilibrium value is identical to the Shannon entropy of the equilibrium set. Nevertheless, on our time scale, physical laws do create highly ordered or low entropy subsystems (i.e., high D-information structures), provided the subsystem is open. In such cases, new forms of order emerge by repackaging existing order, or by accessing stored energy to create new ordered structures while at the same time ejecting heat and/or disordered waste elsewhere in the

universe. Nevertheless, the overall entropy of the universe increases over time.

There is no point in arguing, as Dembski does (2002b, 173), that an unlikely local entropy decrease can only occur because of an injection of information, unless it can be shown that there is no natural way of generating the local order. So far, the observed order of the earth's biosystem is a consequence of solar energy being used to create more ordered structures while disorder, mainly as heat, is ejected.

## CONCLUSION

There are several serious flaws with Dembski's claim that an explanatory filter can be used to provide clear evidence that structures observed in the universe require a design explanation outside of nature. In summary:

- Dembski's design template eliminates natural causes too early, thereby forcing a design explanation when none is warranted. The choice is not between chance and ID, but between natural laws and ID.
- Dembski's attempt to define an information measure, CSI to identify ordered structures is inconsistent. A modified measure based on Kolmogorov's deficiency in randomness is a much more consistent and useful measure of order.
- Dembski's randomness test is too ambiguous and unclear to be workable, as is demonstrated by the fact it has not been credibly applied to any realistic situation. Moreover, as any workable randomness test can be expressed in terms of a universal Martin-Löf randomness test (Li and Vitányi 2008, 134), even if the Dembski test could be made operational in a convincing manner, it is simpler and more convenient to replace the Dembski test by a universal Martin-Löf test based on Kolmogorov's deficiency in randomness.
- Even if Dembski's information measure is modified to be made consistent, the modified measure, defined here makes it clear that the law of conservation of information is no more than the second law of thermodynamics. Dembski seems to need a new law to justify the injection of external order into the universe for ideological reasons.

While the questions Dembski raises are worth considering, a better question is to ask where the initial order in the universe came from, rather than searching for the injection of order in an evolving universe. Furthermore, there is no evidence that Dembski's approach offers anything from a scientific point of view, as the universal randomness test, and the more rational

decision process considered here does it all. As a consequence, the scientific community should only engage in any discussion on the possibility of design interventions in nature if the discussion is articulated in terms of AIT. Discussion on any other basis will achieve little.

## REFERENCES

- Behe, Michael. 1996. *Darwin's Black Box*. New York, NY: The Free Press.
- Bennett, Charles H. 1982. "Thermodynamics of Computation—A Review." *International Journal of Theoretical Physics* 21:905–40.
- . 1987. "Demons, Engines, and the Second Law." *Scientific American* 257:108–16.
- Chaitin, Gregory J. 1966. "On the Length of Programs for Computing Finite Binary Sequences." *Journal of the ACM* 13:547–69.
- . 1975. "A Theory of Program Size Formally Identical to Information Theory." *Journal of the ACM* 22:329–40.
- Dembski, William A. 1998. *The Design Inference: Eliminating Chance through Small Probabilities*. New York, NY: Cambridge University Press.
- . 2002a. "Intelligent Design as a Theory of Information." Available at [http://www.arn.org/docs/dembski/wd\\_idtheory.htm](http://www.arn.org/docs/dembski/wd_idtheory.htm).
- . 2002b. *No Free Lunch: Why Specified Complexity Cannot Be Purchased without Intelligence*. Lanham, MD: Rowman & Littlefield.
- . 2005. "Specification: The Pattern that Signifies Intelligence." Available at <http://www.designinference.com/documents/2005.06.Specification.pdf>.
- Discovery Institute. 2012. "Discovery Institute Center for Science and Culture: Intelligent Design." Retrieved August 2012, from <http://www.intelligentdesign.org/whatisid.php>.
- Elsberry, Wesley, and Jeffrey Shallit. 2011. "Information Theory, Evolutionary Computation, and Dembski's Complex Specified Information." *Synthese* 178:237–70.
- Gács, Péter. 2010. "Lecture Notes on Descriptive Complexity and Randomness." Available at <http://www.cs.bu.edu/~gacs/papers/ait-notes.pdf>.
- Jones, Dan. 2008. "Engines of Evolution." *New Scientist* 197(2643):40–43.
- Kolmogorov, Andrey N. 1965. "Three Approaches to the Quantitative Definition of Information." *Problems of Information Transmission* 1:1–7.
- Landauer, Rolf W. 1961. "Irreversibility and Heat Generation in the Computing Process." *IBM Journal of Research and Development* 5:183–91.
- Leff, Harvey S., and Andrew F. Rex. 1990. *Maxwell's Demon: Entropy, Information, Computing*. Princeton, NJ: Princeton University Press.
- Levin, Leonard. 1974. "Laws of Information Conservation (Nongrowth) and Aspects of the Foundation of Probability Theory." *Problems of Information Transmission* 10:206–10.
- Li, Ming, and Paul M. B. Vitányi. 2008. *An Introduction to Kolmogorov Complexity and Its Applications*, 3rd. ed. New York, NY: Springer-Verlag.
- Lloyd, Seth. 2002. "Computational Capacity of the Universe." *Physical Review Letters* 88:237901–4.
- Martin-Löf, Per E. R. 1966. "The Definition of Random Sequences." *Information and Control* 9:602–19.
- Miller, Kenneth. R. 2004. "The Flagellum Unspun: The Collapse of 'Irreducible Complexity'." In *Debating Design: From Darwin to DNA*, eds. William A. Dembski and Michael Ruse, 81–97. Cambridge, UK: Cambridge University Press.
- Musgrave, Ian. 2006. "Evolution of the Bacterial Flagellum." In *Why Intelligent Design Fails: A Scientific Critique of the New Creationism*, ed. Matt Young and Taner Edis, 72–84. New Brunswick, NJ: Rutgers University Press.
- Shallit, Jeffrey, and Wesley Elsberry. 2004. "Playing Games with Probability: Dembski's Complex Specified Information." In *Why Intelligent Design Fails: A Scientific Critique of the New Creationism*, ed. Matt Young and Taner Edis, 123–38. New Brunswick, NJ: Rutgers University Press.
- Shannon, Claude E., and Warren Weaver. 1949. *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

- Solomonoff, Ray J. 1964. "A Formal Theory of Inductive Inference." *Information and Control* 7:1–22 (part 1), 224–254 (part 2).
- Van Till, Howard J. 2002. "E. Coli at the No Free Lunchroom: Bacterial Flagella and Dembski's Case for Intelligent Design." Posted AAAS website, DoSER section. Available at <http://www.metanexus.net/essay/e-coli-no-free-lunchroom>.
- . 2013. Private Communication.