DEMBSKI'S SPECIFIED COMPLEXITY: A SIMPLE ERROR IN ARITHMETIC

HOWARD A. LANDMAN

Abstract. We show that the derivation of Dembski's formula for specified complexity contains a simple but enormous error, causing it to be wrong by a factor of 10^{117} or so. Correcting the derivation leads to a formula which, while more valid, has rather unsurprising implications.

Contents

1. Introduction

William Dembski, a proponent of Intelligent Design, has published several books and articles arguing that it is possible to determine, on mathematical grounds, when a given object or event can reasonably be inferred to have been designed by an intelligent agent rather than having occurred by some other means. He proposes a measure called *specified complexity* or *complex specified information*, and gives a formula for calculating it. The most recent and complete presentation of this idea is his paper *Specification: The Pattern That Signifies Intelligence*[1], which is available on Dembski's own website.

There has been a great deal of criticism of this idea on philosophical, logical, statistical, and other grounds. While we agree with much of this criticism, in this paper we wish to focus on the fact that Dembski

Date: July 27, 2008.

makes an elementary blunder in the derivation of his formula, so that the published formula is enormously wrong.

In what follows, we first present Dembski's derivation of the formula, stripped of any extraneous baggage. We then give two lines of analysis which both point to the error; one based on information theory, and another based on dimensional analysis. Finally, we consider a corrected derivation and formula, and discuss its meaning and implications.

2. Dembski's Derivation

Dembski starts with the idea of what he calls the *chance hypothesis* **H**. This is, roughly, the idea that the event in question happened "randomly". He is rather vague on the precise definition, and so it is possible that his entire theory doesn't apply at all to processes such as biological evolution which have both chance (e.g. mutation) and non-chance (e.g. selection) components. It may not even apply to any process with forces (e.g. water running downhill) or any process with iteration (e.g. a computer program with loops). However, if we are generous in our interpretation, we could interpret **H** as being equivalent to the *maximum entropy principle* which is widely used in physics and which most scientists would not find objectionable.

Different situations have different probabilities, so the next step is to explicitly name the probability distribution $\mathbf{P}(\cdot \mid \mathbf{H})$. This gives the probability that some (unspecified) event will occur under the hypothesis **H**. For a specific event E we would use $\mathbf{P}(E \mid \mathbf{H})$, and for a target pattern T which would be exhibited by some events and not others, $\mathbf{P}(T \mid \mathbf{H})$. There is really no mathematical difference between "event" and "pattern" here; a pattern is just a kind of event. $\mathbf{P}(T \mid \mathbf{H})$ should be read as "the probability of T, given **H**", and is a number between 0 and 1. So far so good.

Next Dembski introduces the idea of an observing agent with the capability to recognize patterns. No real definition of the agent is given, but the intent is to adjust for the fact that there will usually be very many possible patterns which could be recognized, and that this makes the chance that *some* agent would recognize *some* pattern much higher than the chance for any single pattern T alone. He defines $\varphi_S(T) = |\{U \in patterns(\Omega) | \varphi'_S(U) \leq \varphi'_S(T)\}|$ as "the number of patterns for which Ss semiotic description of them is at least as simple as Ss semiotic description of T". Dembski acknowledges that this makes $\varphi_S(T)$ subjective; different agents S may have different description methods and hence calculate different values for $\varphi_S(T)$. However, this is a relatively small problem and need not concern us here. He concludes: "Thus, if $\varphi_S(T) = n$, there are at most n patterns whose descriptive complexity for S does not exceed that of T. $\varphi_S(T)$ defines the *specificational resources* that S associates with the pattern T."

Multiplying these two together gives the product $\varphi_S(T) \cdot \mathbf{P}(T \mid \mathbf{H})$ which, Dembski says, "provides an upper bound on the probability (with respect to the chance hypothesis $\mathbf{H}$) for the chance occurrence of an event that matches any pattern whose descriptive complexity is no more than T and whose probability is no more than $\mathbf{P}(T \mid \mathbf{H})$". This is, unfortunately, incorrect; just knowing that complexity$(\tilde{T}) \leq$ complexity(T) does not guarantee that $\mathbf{P}(\tilde{T} \mid \mathbf{H}) \leq \mathbf{P}(T \mid \mathbf{H})$. Simpler patterns do not necessarily have lower probabilities of occurrence. The derivation breaks down here.

This is an important point, so it's worth giving a concrete example. Suppose our space consist of all possible strings of ten decimal digits, and our target pattern T is "contains 3963462349". (This is a string of 10 digits taken from a table of random numbers, so it is unlikely that it can be compressed much.) One simpler pattern $\tilde{T}$ is "contains 1". Yet here $\mathbf{P}(T \mid \mathbf{H})$ is 10^{-10} and $\mathbf{P}(\tilde{T} \mid \mathbf{H})$ is about 0.35, nearly 10^{10} times greater.

This error is, however, fixable. Instead of Dembski's original definition of $\varphi_S(T) = |\{U \in \mathit{patterns}(\Omega) | \varphi'_S(U) \leq \varphi'_S(T)\}|$, we can define $\hat{\varphi}_S(T) = |\{U \in \mathit{patterns}(\Omega) | (\varphi'_S(U) \leq \varphi'_S(T)) \wedge (\mathbf{P}(U \mid \mathbf{H}) \leq \mathbf{P}(T \mid \mathbf{H}))\}|$, i.e. "the number of patterns for which Ss semiotic description of them is at least as simple as Ss semiotic description of T and also their probability is no greater than that of T". With this redefinition, the formula $\hat{\varphi}_S(T) \cdot \mathbf{P}(T \mid \mathbf{H})$ becomes a valid upper bound, so a modified derivation can proceed.

To convert this probability into a measure of information, which Dembski calls *specificity*, we take minus the logarithm base 2 to get it in bits: $\sigma = -\log_2(\hat{\varphi}_S(T) \cdot \mathbf{P}(T \mid \mathbf{H}))$. This is a standard operation and needs no comment.

Dembski then considers two additional factors, M, the number of agents that might be witnessing events, and N, "the number of opportunities for such events to happen". It does not seem necessary to include M in the analysis of a single agent making a single observation and decision; it appears that Dembski is trying to consider the case where M agents are making decisions and he wants to make sure that *none* of them are in error. This subtly redefines the problem, but since M adds conservatism, it does not invalidate the derived inequalities. In fact, Dembski later merges $M{\cdot}N$ and replaces their product with a single number, so this issue goes away.

After incorporating M and N, we have Dembski's *context-dependent specified complexity*, $\tilde{\chi} = \log_2(M{\cdot}N \cdot \hat{\varphi}_S(T) \cdot \mathbf{P}(T \mid \mathbf{H}))$.

Dembski then notes that "As defined, $\tilde{\chi}$ is context sensitive, tied to the background knowledge of a semiotic agent S and to the context of inquiry within which S operates. Even so, it is possible to define specified complexity so that it is not context sensitive in this way. Theoretical computer scientist Seth Lloyd has shown that 10^{120} constitutes the maximal number of bit operations that the known, observable universe could have performed throughout its entire multi-billion year history. This number sets an upper limit on the number of agents that can be embodied in the universe and the number of events that, in principle, they can observe. Accordingly, for any context of inquiry in which S might be endeavoring to determine whether an event that conforms to a pattern T happened by chance, $M{\cdot}N$ will be bounded above by 10^{120}." This leads to his final formula $\chi = -\log_2(10^{120} \cdot \hat{\varphi}_S(T) \cdot \mathbf{P}(T \mid \mathbf{H}))$, which is claimed to be universally valid.

Dembski's conclusion is that "if $10^{120} \cdot \hat{\varphi}_S(T) \cdot \mathbf{P}(T \mid \mathbf{H}) < \frac{1}{2}$ or, equivalently, that if $\chi = -\log_2(10^{120} \cdot \hat{\varphi}_S(T) \cdot \mathbf{P}(T \mid \mathbf{H})) > 1$, then it is less likely than not on the scale of the whole universe, with all replicational and specificational resources factored in, that E should have occurred according to the chance hypothesis $\mathbf{H}$."

3. The Error

Hidden in this last step is a subtle error with enormous consequences. What Lloyd[2] actually hypothesizes is that "The universe can have performed no more than 10^{120} ops on 10^{90} bits." The number of bits 10^{90} is "an amount of information equal to the logarithm of its number of accessible states", so the number of accessible states is $2^{10^{90}}$. It is this number which bounds the replicational resources, rather than 10^{120}. Dembski has, in effect, taken the logarithm twice when he should have taken it only once.

It is worth noting that this is consistent with Dembski's own definition of resources. Compare, for example, how he computes the specificational resources for the bacterial flagellum. We are merely computing the replicational resources the same way, as the number of total possibilities.

If we make this substition, then the design criterion changes from $10^{120} \cdot \hat{\varphi}_S(T) \cdot \mathbf{P}(T \mid \mathbf{H}) < \frac{1}{2}$ to $2^{10^{90}} \cdot \hat{\varphi}_S(T) \cdot \mathbf{P}(T \mid \mathbf{H}) < \frac{1}{2}$, a shift which renders it useless. It basically says we should infer non-randomness if we see any event or pattern that is more unlikely than the entire state of the universe. But clearly, no subset or partial view of the universe can be more unlikely than the universe as a whole. (This conclusion is trivial, and can be proven quite easily without going through Dembski's elaborate constructions. Simply note that the probability of the rest of the universe, independent of the part under consideration, cannot be greater than 1.) Thus it will never be possible to infer non-randomness by this method.

We shall now consider two lines of supporting analysis.

3.1. **Information-Theoretic Analysis.** There is not space here to give a complete introduction to information theory. A few good starting places are Shannon's classic paper[3], Tom Schneider's primer[4], and chapter 4 of John Avery's *Information Theory and Evolution*[5].

Information is always a reduction in *uncertainty*, also called *informational entropy* (to distinguish it from physical entropy, which is proportional but has different units; see [5]). That is, $I = U(\text{before}) - U(\text{after})$. Thus information and uncertainty have the same units (e.g. bits) but opposite sign; information is negative uncertainty, uncertainty is negative information.

U can either be calculated as the logarithm of the number of states (a la Boltzmann) or the negative logarithm of the probability of a single state (a la Shannon); in general, if not all probabilities are equal, then Shannon's formula $U = \sum_i -p_i \log_2 p_i$ must be used. Shannon's theorem proves that, given a few simple assumptions, this is the *only* formula which can work.

Dembski is attempting to calculate a kind of uncertainty using something like Shannon's formula. To do this, he needs to take the logarithm of the number of possibilities. But Lloyd's 10^{120} bit-operations and 10^{90} bits are *already* the logarithms of the number of possible computational sequences and the number of possible register states respectively. To insert either of them inside the logarithm in his formula is a mistake; what needs to go there is the thing (sequences or states) of which they are the logarithm.

3.2. **Dimensional Analysis.** *Dimensional Analysis* (DA) is the art of keeping your units straight. For example, suppose we know that a car is traveling 60 miles per hour, but we want to know its speed in feet per second. We know there are 5280 feet in a mile, and 60 minutes in an hour, and 60 seconds in a minute; how do we combine all these to get the correct answer?

DA treats units as if they were unknown algebraic quantities, and tells us to arrange them so that they can be cancelled out. So $60\frac{\text{miles}}{\text{hour}} \cdot$

$\frac{1}{60}\frac{\text{hour}}{\text{minutes}} \cdot \frac{1}{60}\frac{\text{minute}}{\text{seconds}} \cdot 5280\frac{\text{feet}}{\text{mile}} = 88\frac{\text{feet}}{\text{second}}$; the miles, hours, and minutes cancel out.

It is not not legal in DA to take the logarithm of a dimensional unit, or indeed any dimensioned quantity. Because the log of a product is the sum of the logs, we would have to have e.g. $\log(88 \text{ feet}) = \log(88) + \log(\text{foot})$; but "the log of foot" is an undefined quantity. Thus, the argument to any logarithm must be dimensionless. Probabilities are dimensionless, as are pure numbers.

If we examine the units in Dembski's derivation, at first everything goes well, $\varphi_S(T)$ has unit "patterns", and in context $\mathbf{P}(T \mid \mathbf{H})$ can be considered to have units "probability per pattern", so that $\varphi_S(T) \cdot \mathbf{P}(T \mid \mathbf{H})$ ends up as a dimensionless probability, Thus Dembski's specificity $\sigma = -\log_2(\varphi_S(T) \cdot \mathbf{P}(T \mid \mathbf{H}))$ is legal and has unit"bits".

However, Lloyd's 10^{120} has dimension "bit-operations" (and his 10^{90} has dimension "bits"). Neither of these can be put inside the logarithm without something else to cancel out the units. Since there is nothing else, Dembski's $10^{120} \cdot \hat{\varphi}_S(T) \cdot \mathbf{P}(T \mid \mathbf{H})$ must have dimension bit-operations and hence taking the log of it is illegal.

4. Summary

We have shown that Dembski's derivation of his formula $\chi = -\log_2(10^{120} \cdot \varphi_S(T) \cdot \mathbf{P}(T \mid \mathbf{H}))$ contains (at least) two mistakes, one of them small and repairable without changing the argument much, and the other large and fundamental. Analysis indicates that a correct derivation would give $2^{10^{90}}$ instead of 10^{120} as the leading factor. However, the corrected formula is useless for Dembski's stated purposes and does not support any of his philosophical conclusions.

References

[1] William A. Dembski, "Specification: The Pattern That Signifies Intelligence", version 1.22, August 15, 2005.

[2] Seth Lloyd, "Computational capacity of the universe", arXiv:quant-ph/0110141v1, October 24 2001.

[3] C. E. Shannon, "A mathematical theory of communication", *Bell System Technical J.*, 27:379-423, 1948.

[4] Tom Schneider, "Information Theory Primer With an Appendix on Logarithms", February 20 2007. www.lecb.ncifcrf.gov/~toms/paper/primer/

[5] John Avery, *Information Theory and Evolution*, World Scientific 2003.

520 N. Sherwood St. #24, Fort Collins, Colorado 80521

E-mail address: howard@riverrock.org