The Conservation of Information: Measuring the Cost of Successful Search

William A. Dembski

Center for Science and Culture
Discovery Institute
Seattle, WA 98101

dembski@discovery.org

version 1.1, 6 May 2006

Abstract

Many spaces that need to be searched in the sciences are too unwieldy for random search to stand any hope of success. Success instead requires a nonrandom search. But how does one *find* a nonrandom search that stands a good chance of success? Even to pose the question this way suggests that such nonrandom searches do not magically materialize but need themselves to be discovered by a process of search. The question then naturally arises whether such a "search for a search" is any easier than the original search. This paper establishes a conservation of information theorem according to which the information required to find a successful search is always at least as large as the information required to successfully complete the original search. This result shows that information, like money, obeys strict accounting principles, leaves a trail, and can only originate from a prior information source.

1 Random Search

Successful search always incurs an information cost. The purpose of this brief paper is to characterize, as an accountant might, the lower limit below which that cost cannot be reduced. The baseline against which to calculate that cost is random search. In a random search, individual elements from a search space are independently sampled with respect to a given probability distribution. Thus, for a sample of size 1, the probability of success will be some value $p > 0$, and for a sample of size n the probability of success will be $1 - (1 - p)^n$.

To derive this last term, take independent and identically distributed random variables $X_1, X_2,..., X_n$ that have probability p of landing in some set A. In that case, the probability of successfully landing in A is

$$\begin{aligned}
\mathbf{P}(X_1 &\in A \text{ or } X_2 \in A \text{ or } \cdots \text{ or } X_n \in A) \\
&= \sum_{i=1}^{n} \mathbf{P}(X_1 \notin A \text{ and } X_2 \notin A \text{ and } \cdots \text{ and } X_{i-1} \notin A \text{ and } X_i \in A) \\
&= \sum_{i=1}^{n} \mathbf{P}(X_1 \notin A)\mathbf{P}(X_2 \notin A)\cdots\mathbf{P}(X_{i-1} \notin A)\mathbf{P}(X_i \in A) \qquad \text{[by indep.]} \\
&= \sum_{i=1}^{n} (1-p)^{i-1} p \\
&= 1 - (1-p)^n.
\end{aligned}$$

In the known physical universe, the number of elements that can be sampled from a search space is always strictly limited. At the time of this writing, the fastest computer is the Department of Energy's IBM BlueGene/L with over 130,000 processors and peaking at 367 teraflops.[1] If we imagine each floating point operation as able to take a sample of size 1, then this computer, even when run over the duration of the physical universe (i.e., 12 billion years), would be able to sample at most $m = 10^{33}$ elements from the search space.

Seth Lloyd (2002) has shown that 10^{120} is the maximal number of bit operations that the known, observable universe could have performed throughout its entire multi-billion year history. Thus, 10^{120} is an absolute limit on the sample size of any search. In the sequel, we will treat m as the upper limit on the number of elements that a given search can sample.

Given that a random search has probability of success $p_n = 1 - (1-p)^n$, it follows that as n gets arbitrarily large, p_n converges to 1. Thus, with unlimited sample size, random search becomes a perfect search. But sample size, as we just noted, is always limited by m. Moreover, for virtually all interesting problems, p tends to be so small that $p_m = 1-(1-p)^m$ tends also to be very small (typically, p is many orders of magnitude less than $1/m$ so that by Taylor's expansion p_m is approximately mp, which is then still very small).

To be successful, a search will therefore need to do better than random search. Such a nonrandom search will have probability q_n of success for a sample of size n (note that both p_n and q_n are monotonically increasing in n). But since m is an upper limit on the number of elements that can be sampled from a search space (whether randomly or nonrandomly), q_m needs to be reasonably close to 1 if the nonrandom search is to stand a good chance of success. Let us next turn to the information cost associated with such searches.

[1] See http://www.top500.org/lists/2005/11/basic (last accessed March 8, 2006).

2 Added Information

Binary codes provide the simplest and most cost-efficient way of handling information (in particular, such codes use the least memory and bandwidth).[2] Hence, the most convenient way for information theorists to measure information is in bits. For this reason, the logarithm to the base 2 has become the canonical logarithm for information theorists. Given an event A of probability p, the *information* associated with A is therefore defined as

$$I(A) =_{def} -\log_2 p.^3$$

Consider now the random search S with probability p of success for a single query (i.e., for a sample of size 1). Let S_n denote a sample of size n for this search. In that case, the information associated with the success of this search is $I(S_n) = -\log_2 p_n$, where $p_n = 1 - (1-p)^n$. Since we assume that n is always bounded above by m and that p is many orders of magnitude less than $1/m$ (implying that $1 - (1-p)^n \approx np$), it follows that $I(S_n) \approx -\log_2 np$. If we now let T denote a nonrandom search with probability of success q_n for a sample of size n, then the information associated with the success of this search is $I(T_n) = -\log_2 q_n$.

Since p_n and q_n are monotonically increasing in n, the information associated with these searches (i.e., $I(S_n)$ and $I(T_n)$) goes down as n increases. This seems counterintuitive because we tend to think that the larger the sample size of a search, the better our chance of success and therefore the more information gets generated. This intuition is correct, but the information measure just described does not capture it. Rather, another type of information measure is needed to capture it, namely, what may be called the *added information*. Given events A and B of probability p and q respectively, the information that B adds to A is defined as

$$I_+(A:B) =_{def} -\log_2 p + \log_2 q = \log_2 q/p.^4$$

An immediate consequence of this definition is that A adds no information to itself, i.e., $I_+(A:A) = 0$. Another consequence is that because q, as a probability, cannot exceed 1, $I_+(A:B)$ can never exceed $I(A) = -\log_2 p$. A third

[2] For the privileged place of the binary code, see von Baeyer (2004: 30–31).

[3] Information theorists sometimes refer to the definition of information just given as the *surprisal* associated with a particular event (the smaller the event's probability, the bigger the "surprise" associated with its occurrence—see Dretske 1981: 10). Given events $A_1, A_2, ..., A_m$ that are mutually exclusive and exhaustive, and given that the probability of A_i is p_i ($1 \leq i \leq m$, $p_1 + p_2 + \cdots + p_m = 1$, no $p_i = 0$), information theorists define a more general information measure, known as *entropy*:

$$H =_{def} -\sum_{i=1}^{m} p_i \log_2 p_i.$$

For our purposes, the surprisal is all we need.

[4] Added information as defined here is the non-averaged form of what information theorists call the *relative entropy* (also known as the Kullback-Leibler distance). See Cover and Thomas (1991: 18) as well as the previous note.

consequence is that because q can approach 0, $I_+(A : B)$ can assume arbitrarily large negative values (thus enabling added information to model "stupid" searches, which do worse than random search).

For the random search S, it now follows that increasing the sample size adds information. Because $1 - (1 - p)^n \approx np$, it follows that

$$I_+(S_1 : S_n) = \log_2[1 - (1 - p)^n]/p \approx \log_2 n.$$

This result makes good intuitive sense, indicating that increasing the sample size of a random search by n adds only $\log_2 n$ bits of information to the search. Compare this to a nonrandom search T that for each query (i.e., each sample of size 1) reduces the search space in half and thus removes half the uncertainty (as in interval halving). In that case, the probability of a successful search with a sample of size n is $q_n = 2^{n-1}p$. It follows that

$$I_+(T_1 : T_n) = \log_2 2^{n-1}p/p = n - 1.$$

In this case, increasing the sample size of the search by n also increases the information by n.

Note, however, that this result only holds so long as $2^{n-1}p \leqslant 1$. Accordingly, the maximal amount of information that multiple queries (samples) can add to a search occurs when multiple queries guarantee success, raising the probability of success to 1. Thus, for an arbitrary search U such that a single query has probability p of success and such that n queries guarantee success with probability 1, $I_+(U_1 : U_n) = \log_2 1/p$ is maximal.

Finally, there is no reason to confine added information to queries from a single search. Consider two searches, V and W. It makes good sense to ask the degree to which n queries of search W add information to k queries of search V. In that case, $I_+(V_k : W_n) = \log_2 q/p$ where q is the probability of success in n queries of W and p is the probability of success in k queries of V. This formula is well-defined mathematically. In practice, however, it will be significant only if both searches (V and W) employ the same criterion of success (e.g., two biological searches where success means locating the same family of functional proteins).

3 The Cost of Success

Let us now return to the main point of interest, namely, how a nonrandom search T adds information to a random search S. As before, suppose S has probability p of success for a single query and thus probability $p_n = 1 - (1 - p)^n$ of success for n queries. Hence, $I(S_1) = -\log_2 p$ and $I(S_n) = -\log_2 p_n \approx -\log_2 np$. This last (approximate) equality holds for $n \leqslant m$ provided that m is the maximal number of queries (i.e., maximal sample size) and provided that p is, as we are assuming, many orders of magnitude less than $1/m$.

Next consider the nonrandom search T. Given that T has probability of success q_n for n queries, $I(T_n) = -\log_2 q_n$. Since m is the maximal number of queries, we focus on the added information $I_+(S_m : T_m) = \log_2 q_m/p_m$, which calculates the amount of information that T_m adds to S_m. We assume that both searches, S and T, employ the same criterion of success and that T is significantly more effective at conducting a successful search than S (i.e., $q_m \gg p_m$).

Since an m-query search is equivalent to a 1-query search consisting of an m-query block, we simplify our notation and drop all m-subscripts. Accordingly, S_m becomes, S, T_m becomes T, q_m becomes q, p_m becomes p, and $I_+(S_m : T_m)$ becomes $I_+(S : T)$.

A further simplification will now be useful in the sequel. By the previous simplification, S and T are one-query searches with probabilities p and q respectively of success where $0 < p \ll q \leqslant 1$. Here q is so much bigger than p that T's success is largely assured whereas S is highly unlikely to be successful. Now, because S is a random search, if $q < 1$, it could happen that S succeeds but T fails. Mathematically, however, it ends up being easier to fold S's successes into T's successes. This is easily accomplished by redefining T as the search that succeeds whenever either the original T or S or both succeed.[5] Doing so entails no loss of generality since S is the baseline against which the success of T is gauged.

If T now has this property (i.e., a success for S entails a success for T), we can consider the modified searches S^* and T^* where, in sampling from the underlying search space, one ignores the cases where T^* does not succeed. In other words, S^* and T^* are formed from S and T by rejection sampling, i.e., by rejecting those samples of T (and thus the corresponding samples of S) for which the T-search fails.

When this is done, it immediately follows that T^* succeeds with probability 1 and S^* succeeds with probability $\frac{p}{q}$, which is strictly greater than p since q is here assumed to be strictly less than 1. Accordingly, for the modified searches S^* and T^*,

$$I_+(S^* : T^*) = -\log_2 \tfrac{p}{q} + \log_2 1 = \log_2 q/p,$$

which is just the original added information $I_+(S : T)$. The bottom line is that in analyzing added information of searches, it is enough to consider searches that guarantee success with probability one in relation to random searches with a suitable probability of success (for S^* that probability of success is $\frac{p}{q}$).

In the sequel, therefore, if we rewrite S for S^* and T for T^*, we can limit ourselves to one-query searches for which the probability of success of the ran-

[5] It could be that T succeeds whenever S does. In that case, there is no need to redefine T. On the other hand, it could be that they don't. In that case, q, the probability that a redefined T succeeds, will need to be adjusted up. For instance, suppose that S and T are stochastically independent searches. In that case, let A represent success for S on a given query and let B represent success for T on that same query. Then the probability that S or T (or both) succeed on a given query is given by $\mathbf{P}(A \text{ or } B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \text{ and } B)$. But $\mathbf{P}(A) = p$, $\mathbf{P}(B) = q$, and by independence $\mathbf{P}(A \text{ and } B) = \mathbf{P}(A) \times \mathbf{P}(B) = pq$. Accordingly, $\mathbf{P}(A \text{ or } B) = p + q - pq = q + p(1 - q)$, which is the new, slightly bigger q-value for the redefined T.

dom search S is p (formerly $\frac{p}{q}$) and the probability of success of the nonrandom search T is 1.

4 The Search for a Search

The key question that now needs to be answered is, What is the source of the information that makes T a more effective search than S? Where does this information come from? Short of magically materializing, this information must itself derive from a search. But since this information is already embedded in a search, namely T, it must derive from searching a collection of searches in which both S and T reside.

To see what's at stake, imagine that you are on an island with buried treasure. The island is so large that a random search (S) is highly unlikely to succeed. Fortunately, you have a treasure map that will guide you to the treasure with unfailing accuracy. In other words, the treasure map allows you to perform a nonrandom search (T) that has probability 1 of success.

But where did you get the treasure map? Treasure maps reside in a larger collection of all conceivable treasure maps. The vast majority of these maps will not lead to the treasure. How then did you happen to find the right one among all conceivable treasure maps? What special information did you have so that you could find the right map?

Given that the nonrandom search T is itself the result of a search (i.e., a search for a search), we can now perform the same analysis as we did on the original search space. Accordingly, just as the original search required a baseline random search, so we need to consider a baseline random search over this higher-level space of searches. Denote such a random search by $\overline{S}_1, \overline{S}_2, \overline{S}_3$,... where the $\overline{S}_i$s are independent and identically distributed taking values in the space of searches to which S and T belong. This higher-level random search $\overline{S}$ will be consistent with the original random search S, which serves as the baseline for searching the original space, provided that, on average, the $\overline{S}_i$s yield a search whose probability of success on the original space is no greater than p.

A few further simplifications are now in order: let the searches S and T represent success and failure on the binary set $\{0, 1\}$, 0 for failure, 1 for success. For S the probability of success is p and for T it is 1. It follows that we can, without loss of generality, represent the searches S and T as numbers (i.e., probabilities) from the unit interval $[0, 1]$, S as p and T as 1.

Accordingly, as with lower-level search, higher-level search can now be represented as a probability distribution, though this time on the unit interval $[0, 1]$ instead of on the binary set $\{0, 1\}$. Moreover, since the $\overline{S}_i$s yield a search on the original space that, on average, has probability no greater than p of success, for the random search $\overline{S}_1, \overline{S}_2, \overline{S}_3$,..., the $\overline{S}_i$s follow some probability distribution μ on $[0, 1]$ such that the mean of μ does not exceed p, i.e., $\int_0^1 x d\mu(x) \leqslant p$. For many applications, one can think of this integral as $\int_0^1 x f(x) dx$ where f is a probability density function on the unit interval.

Now, for any number n, let n^* denote the number of $\overline{S}_i$s in $\overline{S}_1, \overline{S}_2,..., \overline{S}_n$ which successfully locate the search T. Alternatively, since we are, without loss of generality, representing searches on the original space as probabilities that these searches succeed, n^* denote the number of $\overline{S}_i$s in $\overline{S}_1, \overline{S}_2,..., \overline{S}_n$ that attain the probability 1 (T being a search that succeeds with probability 1).

By the Strong Law of Large Numbers it now follows that $\frac{1}{n}(\overline{S}_1+\overline{S}_2+\cdots+\overline{S}_n)$ converges almost everywhere to some number less than or equal to p (because the $\overline{S}_i$s are μ-distributed and μ has mean/expectation less than or equal to p). Also by the Strong Law of Large Numbers, the stochastic measures $\frac{1}{n}(\delta_{\overline{S}_1} + \delta_{\overline{S}_2} + \cdots + \delta_{\overline{S}_n})$ converge almost everywhere weakly to μ (see Parthasarathy 1967: 52–53; for stochastic measures in general see Kallenberg 1986).[6]

It follows that the limit as n goes to infinity of the fraction $\frac{n^*}{n}$ is the limit as n goes to infinity of the probability that $\frac{1}{n}(\delta_{\overline{S}_1} + \delta_{\overline{S}_2} + \cdots + \delta_{\overline{S}_n})$ assigns to the singleton $\{1\}$ (which, in our representation, corresponds to the searches that succeed with probability 1 on the original space). And this number, because $\frac{1}{n}(\delta_{\overline{S}_1} + \delta_{\overline{S}_2} + \cdots + \delta_{\overline{S}_n})$ converges almost everywhere weakly to μ, is bounded above by $\mu(\{1\})$ (see Billingsley 1999: 16). $\overline{S}$ therefore models a search for a search and is distributed as the probability measure μ. Accordingly, $\mu(\{1\})$ is the probability that $\overline{S}$ successfully locates a search over the original space that has probability 1 of successfully searching the original space.

Now, for all probability measures μ on the unit interval whose mean is less than or equal to p, the maximum probability that μ can assign to $\{1\}$ is p, which is attained for the probability measure $\mu^* = (1-p)\delta_0 + p\delta_1$. More formally, $\sup\{\mu(\{1\}) \mid \int_0^1 x d\mu(x) \leqslant p\}$ is attained for $\mu^* = (1-p)\delta_0 + p\delta_1$, which assigns probability p to $\{1\}$. To see this, note that the probability measures in question need to balance two competing objectives: they require their mean/expectation to fall in the interval $[0, p]$ and they require maximizing the amount of mass concentrated in the singleton $\{1\}$. Since $0 < p < 1$, $\mu^* = (1-p)\delta_0 + p\delta_1$ is the probability measure that most perfectly fits this bill.

Having identified $\overline{S}$ as a random search for a search that locates the original random search S, the next order of business is to identify a nonrandom search for a search, call it $\overline{T}$, that locates the original nonrandom search T. By an earlier simplification, we limit our nonrandom searches to those that succeed with probability 1. Accordingly, T guarantees success of the original search with probability 1 and $\overline{T}$, in turn, guarantees success of the search for this search (i.e., T) with probability 1. But since we are representing $\overline{T}$ as a probability measure ν on the unit interval $[0, 1]$ and since we are representing T as the probability 1 in this interval, for $\overline{T}$ to guarantee the success of finding T with probability 1 means that ν must assign all its probability to the set $\{1\}$, which in turn means that ν is just the point mass δ_1.

It follows that $I_+(\overline{S} : \overline{T})$ equals the negative logarithm of the probability

[6] The delta functions here are point masses. Because the points to which these delta functions assign mass here are themselves given by stochastic searches, the probability measures that arise from taking convex linear combinations of these delta functions are stochastic probability measures, i.e., they are probability measures which vary stochastically in the way they assign probability.

that $\overline{S}$ assigns to $\{1\}$, which, as we have seen, is minimized when $\overline{S}$ is represented by the measure $\mu^* = (1-p)\delta_0 + p\delta_1$. And since μ^* assigns probability p to $\{1\}$, this means that $I_+(\overline{S} : \overline{T}) \geqslant -\log_2 p$ for all higher-level random searches $\overline{S}$ that locate S and all higher-level nonrandom searches $\overline{T}$ that locate T. But $I_+(S : T) = -\log_2 p$. We have thus proven the following theorem.

5 The Conservation of Information Theorem

THEOREM (CONSERVATION OF INFORMATION). Suppose S and T are searches over a given search space, S being a random search with probability p of success in a single query and T being a nonrandom search with probability 1 of success in a single query. Suppose further that $\overline{S}$ and $\overline{T}$ are searches over the space of searches in which S and T reside so that $\overline{S}$ on average locates a search of the original space that with probability no more than p successfully searches the original space and that $\overline{T}$ with probability 1 locates a search of the original space what with probability 1 successfully searches the original space. Then the information that $\overline{T}$ adds to $\overline{S}$ is at least as great as the information that T adds to S, i.e.,

$$I_+(\overline{S} : \overline{T}) \geqslant I_+(S : T).$$

Moreover, by a suitable choice of $\overline{S}$, this inequality becomes an equality.

REMARKS. (1) Earlier we saw that there is no loss of generality assuming that T is a perfect search, i.e., one that succeeds with probability 1. In general, however, T may be less than perfect, with probability q of success where q is much bigger than p but nonetheless strictly less than 1. In that case, the Conservation of Information Theorem still holds with $I_+(S : T) = \log_2 q/p$ serving as a lower bound for $I_+(\overline{S} : \overline{T})$.

(2) This theorem depends on representing searches as probabilities of success. Granted, such a representation loses a lot of information about actual searches. In evolutionary computing, for instance, an actual search requires an initialization, a fitness, an update rule, and a stop criterion. All such information, however, merely adds to the problem of finding a higher-level search that locates a successful lower-level search. By representing searches in terms of probability of success, I'm not merely distilling the essence of these searches but in fact being conservative about the amount of information required to find a higher-level search that locates a successful lower-level search.

In concluding this paper, I want briefly to discuss the significance of the Conservation of Information Theorem. For most interesting search problems, random search is highly unlikely to succeed. For a search to succeed, it therefore needs to be a nonrandom search. Such a search, however, does not float in from nowhere. If it is to be explained at all, it must be explained as the outcome of a search in its own right. But what sort of search could that be?

A random search for such a search is, according to the Conservation of Information Theorem, even less likely to succeed than a random search of the original search space. Moreover, a nonrandom search for such a search, insofar as it succeeds, adds, again according to the Conservation of Information Theory, at least as much information as the original nonrandom search. Obviously, going to still higher-level searches won't resolve the problem since there is a regress, and the problem only gets worse as we go up the search hierarchy:

$$I_+(S:T) \leqslant I_+(\overline{S}:\overline{T}) \leqslant I_+(\overline{\overline{S}}:\overline{\overline{T}}) \leqslant \cdots.$$

The problem of explaining the information in the original nonrandom search T that enables it to succeed is therefore not explained away by a hide-the-pea game in which this information is shuffled off to higher-level searches. Rather, this problem must be dealt with on its own terms. In practice, there are only two options here: the information is created as an act of intelligence or it is the unintelligent (mechanical) outworking of preexisting information.

Either option raises questions about the ultimate source of that information. According to Douglas Robertson (1999), the defining feature of intelligence is its ability to create information. Yet, if an act of intelligence created the information, where did this intelligence come from? Was information in turn required to create it? Very quickly this line of questioning pushes one to an ultimate intelligence that creates all information and yet is created by none (see Dembski 2004: ch. 19, titled "Information ex Nihilo").

On the other hand, if the information is the mechanical outworking of preexisting information, the Conservation of Information Theorem suggests that this preexisting information was at least as great in the past as it is now (this being the information that allows the present search to be successful). But then how do we make sense of the fact (if it is a fact) that the information in the universe was less in the past than it is now? Indeed, our present universe, with everything from star systems to living forms, seems far more information-rich than the universe at the moment of the Big Bang.

Holmes Rolston (1999: 352–353, 357) offers some interesting insights in this respect. Writing on the "genesis of information," he notes,

> There are no humans invisibly present (as an acorn secretly contains an oak) in the primitive eukaryotes, to unfold in a lawlike or programmatic way.... On Earth, there really isn't anything in rocks that suggests the possibility of *Homo sapiens*, much less the American Civil War, or the World Wide Web, and to say that all these possibilities are lurking there, even though nothing we know about rocks or carbon atoms, or electrons and protons suggests this is simply to let possibilities float in from nowhere.... The information (in DNA) is interlocked with an information producer-processor (the organism) that can transcribe, incarnate, metabolize, and reproduce it. All such information once upon a time did not exist but came into place; this is the locus of creativity. Nevertheless, on

> Earth, there is this result during evolutionary history. The result involves significant achievements in cybernetic creativity, essentially incremental gains in information that have been conserved and elaborated over evolutionary history. The know-how, so to speak, to make salt is already in the sodium and chlorine, but the know-how to make hemoglobin molecules and lemurs is not secretly coded in the carbon, hydrogen, and nitrogen.... Can one claim that what did actually manage to happen must always have been either probably probable, or, minimally, improbably possible all along the way? Push this to extremes, as one must do, if one claims that all the possibilities are always there, latent in the dust, latent in the quarks. Such a claim becomes pretty much an act of speculative faith, not in present actualities, since one knows that these events took place, but in past probabilities always being omnipresent.... Unbounded possibilities that one posits ad hoc to whatever one finds has in fact taken place—possibilities of any kind and amount desired in one's metaphysical enthusiasm—can hardly be said to be a scientific hypothesis. This is hardly even a faith claim with sufficient warrant. It is certainly equally credible and more plausible, and no less scientific to hold that new possibility spaces open up en route.

In light of Rolston's remarks, the Conservation of Information Theorem pushes us in either of two directions: (1) We explain the information in the universe as the creative act of an intelligence that needs no information in turn to explain it. (2) We explain the information in the universe as the mechanical outworking of the physical laws and processes by which the universe operates and in which this information has always resided (even if concealed from our eyes).

Rolston's point is that empirical evidence does not support (2). Accordingly, he regards adherence to (2) not as a scientific inference but as an act of speculative faith. On the other hand, he leaves open the possibility that empirical evidence might support (1). The Conservation of Information Theorem provides conceptual space within which to marshal and assess such evidence.

ACKNOWLEDGMENT. I'm indebted to Robert Marks for the notion of added information developed in this paper.

References

[1] Billingsley, Patrick, *Convergence of Probability Measures*, 2nd ed. (New York: Wiley, 1999).

[2] Cover, Thomas M., and Joy A. Thomas, *Elements of Information Theory* (New York: Wiley, 1991).

[3] Dembski, William A., *The Design Revolution: Answering the Toughest Questions about Intelligent Design* (Downers Grove, Ill.: InterVarsity, 2004).

[4] Dretske, Fred, *Knowledge and the Flow of Information* (Cambridge, Mass.: MIT Press, 1981).

[5] Kallenberg, Olav, *Random Measures*, 4th ed. (Berlin: Akademie-Verlag, 1986).

[6] Lloyd, Seth, "Computational Capacity of the Universe," *Physical Review Letters* 88(23) (2002): 7901–4.

[7] Parthasarathy, K. R., *Probability Measures on Metric Spaces* (New York: Academic Press, 1967).

[8] Robertson, Douglas, "Algorithmic Information Theory, Free Will, and the Turing Test," *Complexity* 4(3) (1999): 25–34.

[9] Rolston III, Holmes, *Genes, Genesis and God: Values and Their Origins in Natural and Human History* (Cambridge: Cambridge University Press, 1999).

[10] von Baeyer, Hans Christian, *Information: The New Language of Science* (Cambridge, Mass.: Harvard University Press, 2004).