The Search for a Search: Measuring the Information Cost of Higher Level Search

William A. Dembski* and Robert J. Marks II**

*Center for Science & Culture, Discovery Institute
Seattle, WA 98104, USA
E-mail: Robert.Marks@baylor.edu

**Dept. of Electrical & Computer Engineering, Baylor University
Waco, TX 76798, USA

[Received September 19, 2009; accepted April 1, 2010]

Needle-in-the-haystack problems look for small targets in large spaces. In such cases, blind search stands no hope of success. Conservation of information dictates any search technique will work, on average, as well as blind search. Success requires an assisted search. But whence the assistance required for a search to be successful? To pose the question this way suggests that successful searches do not emerge spontaneously but need themselves to be discovered via a search. The question then naturally arises whether such a higher-level “search for a search” is any easier than the original search. We prove two results: (1) The Horizontal No Free Lunch Theorem, which shows that average relative performance of searches never exceeds unassisted or blind searches, and (2) The Vertical No Free Lunch Theorem, which shows that the difficulty of searching for a successful search increases exponentially with respect to the minimum allowable active information being sought.

Keywords: No Free Lunch Theorems, active information, active entropy, assisted search, endogenous information

1. Introduction

Conservation of information theorems [1–3], especially the No Free Lunch Theorems (NFLT’s) [4–8], show that without prior information about a search environment or the target sought, one search strategy is, on average, as good as any other [9]. This is the result of the Horizontal NFLT presented in Section 3.2.

A search’s difficulty can be measured by its endogenous information [1, 10–14] defined as

$$I_\Omega = -\log_2 p$$

where $p$ is the probability of a success from a random query [1]. When there is knowledge about the target location or search space structure, the degree to which the search is improved is determined by the resulting active information [1, 10–14]. Even moderately sized searches are virtually certain to fail in the absence of knowledge about the target location or the search space structure. Knowledge concerning membership of the search problem in a structured class [15], for example, can constitute search space structure information [16].

Endogenous and active information together allow for a precise characterization of the conservation of information. The average active information, or active entropy, of an unassisted search is zero when no assumption is made concerning the search target. If any assumption is made concerning the target, the active entropy becomes negative. This is the Horizontal NFLT presented in Section 3.2. It states that an arbitrary search space structure will, on average, result in a worse search than assuming nothing and simply performing an unassisted search.

The measure of endogenous and active information can also be applied in a meta sense to a Search for a Search (S4S). As one might expect, no active information in the S4S translates to zero active information in the lower-level search (which means that, on average, the lower-level search so found cannot do better than an unassisted search). This result holds for still higher level searches such as a “search for a search for a search” and so on. Thus, without active information introduced somewhere in the search hierarchy, none will be available for the original search. If, on the other hand, active information is introduced anywhere in the hierarchy, it projects onto the original search space as active information.

The target of a S4S is a search algorithm that equals or exceeds a minimally acceptable active information threshold. A higher-level S4S will, itself, have a difficulty as measured by the S4S’s endogenous information. How much? Previous results have shown the difficulty of a S4S as measured by its endogenous information is lower bounded by the desired active information of the target search [14]. We establish a much more powerful result. According to the Vertical NFLT introduced in Section 4.3, the difficulty of a S4S under loose conditions, as measured by the S4S endogenous information, increases exponentially with respect to the active information threshold required in the lower-level search space.
2. Information in Search

All but the most trivial searches require information about the search environment (e.g., smooth landscapes) or target location (e.g., fitness measures) if they are to be successful. Conservation of information theorems [2–5] show that one search algorithm will, on average, perform as well as any other and thus that no search algorithm will, on average, outperform an unassisted, or blind, search. But clearly, many of the searches that arise in practice do outperform blind unassisted search. How, then, do such searches arise and where do they obtain the information that enables them to be successful?

2.1. Blind and Assisted Queries and Searches

Let \( T_1 \in \Omega_1 \) denote a target in a search space, \( \Omega_1 \), so that for uniformity the probability of success, \( p \), is ratio of the cardinality of \( T_1 \) to that of \( \Omega_1 \). The probability, \( p \), is the chance of obtaining an element in the target with a single query assuming a uniform distribution. If nothing is known about the search space or target location, the uniformity of the distribution follows from Bernoulli’s Principle of Insufficient Reason (PrOIR) [14, 17, 18]. Bernoulli’s PrOIR states that in the absence of any prior information, “we must assume that the events ... have equal probability” [17, 18]. Uniformity is equivalent to the assumption that the search space is at maximum informational entropy.

Consider \( Q \) queries (samples) from \( \Omega_1 \) without replacement. Such searches can be construed as a single query when the search space is appropriately defined. For a fixed \( Q \), the augmented search space, \( \Omega_Q \), consists of all sequences of length \( Q \) chosen from \( \Omega_1 \) where no element appears more than once. Note that, appropriately, the original search space, \( \Omega_1 \), is \( \Omega_Q \) for \( Q = 1 \). In general

\[
|\Omega_Q| = \frac{\Omega_1!}{(\Omega_1 - Q)!}.
\]  

The requirement of sampling without replacement requires \( Q \leq \Omega_1 \). If nothing is known about the location or the target, then, from Bernoulli’s PrOIR, any element of \( \Omega_Q \) is as likely to contain the target as any other element.

\[
p_Q = \frac{|T_Q|}{|\Omega_Q|},
\]

where \( T_Q \in \Omega_Q \) consists of all the elements containing the original target, \( T_1 \).

When, for example, \( T_1 \) consists of a single element in \( \Omega \) (i.e., \( |T_1| = 1 \)) and there is no knowledge about the search space structure or target location, then \( p_1 = 1/|\Omega_1| \) and

\[
|T_Q| = p_Q |\Omega_Q| = \frac{Q}{|\Omega_1|} \Omega_Q = Q p_1.
\]

From Eqs. (2) and (3), we have

\[
Q p_1 = \frac{Q}{|\Omega_1|} \Omega_Q = Q p_1,
\]

Order 1

There appears, at first flush, to be knowledge about the search space \( \Omega_Q \) for \( Q > 1 \) since a sequence in \( \Omega_Q \) has the same chance of containing a target as an element in \( \Omega_Q \) with a perturbation of the same elements. This, however, is knowledge we cannot exploit with a single query to \( \Omega_Q \). Also, if the space is exhaustively contracted to eliminate all elements allowing a perturbation rearrangement to another element, the chance of success in a single query remains the same.

Keeping track of \( Q \) subscripts and exhaustively contracted perturbation search spaces is distracting. As such, let \( \Omega \) denote any search space, like \( \Omega_Q \), wherein Bernoulli’s PrOIR is applicable for a single query. We adopt similar notation for the target subspace, \( T \), and the probability, \( p \), of finding the target using a single query. A single query to a multi-query space can then be considered a search. If a search is chosen randomly from such a multi-query space, we are still preforming a blind (or unassisted) search.

3. Active Information and No Free Lunch

Define an assisted query as any choice from \( \Omega_1 \) that provides more information about the search environment or candidate solutions than a blind search. Gauging the effectiveness of assisted search in relation to blind search is our next task. Random sampling with respect to the uniform probability \( U \) sets a baseline for the effectiveness of blind search. For a finite search space with \( |\Omega| \) elements, the probability of locating the target \( T \) has uniform probability given by Eq. (3) without the subscripts:

\[
p = \frac{|T|}{|\Omega|}.
\]

Let \( q \) denote the probability of success of an assisted search. We assume that we can always do at least as well as uniform random sampling. The question is, how much better can we do? Given a small target \( T \) in \( \Omega \) and probability measures \( U \) and \( \varphi \) characterizing blind and assisted search respectively, assisted search will be more effective than blind search when \( p < q \), as effective if \( p = q \), and less effective if \( p > q \). In other words, an assisted search can be more likely to locate \( T \) than blind search, equally likely, or less likely. If less likely, the assisted search is counterproductive, actually doing worse than blind search. For instance, we might imagine an Easter egg hunt where one is told “warmer” if one is moving away from an egg and “colder” if one is moving toward it. If one acts on this guidance, one will be less likely to find the egg than if one simply did a blind search. Be-
cause the assisted search is in this instance misleading, it actually undermines the success of the search.

An instructive way to characterize the effectiveness of assisted search in relation to blind search is with the likelihood ratio \( \frac{\phi}{\Omega} \). This ratio achieves a maximum of \( \frac{1}{p} \) when \( q = 1 \), in which case the assisted search is guaranteed to succeed. The assisted search is better at locating \( T \) than blind search provided the ratio is bigger than 1, equivalent to blind search when it is equal to 1, and worse than blind search when it is less than 1. Finally, this ratio achieves a minimum value of 0 when \( q = 0 \), indicating that the assisted search is guaranteed to fail.

### 3.1. Active Information

Let \( U \) denote a uniform distribution on \( \Omega \) characteristic of an unassisted search and \( \phi \) the (nonuniform) measure on \( \Omega \) for an assisted search. Let \( U(T) \) and \( \phi(T) \) denote the probability over the target set \( T \in \Omega \). Define the active information of the assisted search as

\[
I_+(\phi|U) := \log_2 \frac{\phi(T)}{U(T)}.
\]

Active information measures the effectiveness of assisted search in relation to blind search using a conventional information measure. It characterizes the amount of information \( \phi \) (representing the assisted search) adds with respect to \( U \) (representing the blind search) in the search for \( T \). Active information therefore quantifies the effectiveness of assisted search against a blind-search baseline. The NFLT dictates that any search without active information will, on average, perform no better than blind search.

The maximum value that active information can attain is \( (I_+)_\text{max} = -\log_2 \frac{p}{q} = I_\Omega \), indicating an assisted search guaranteed to succeed (i.e., a perfect search) [1]; and the minimum it can attain is \( (I_+)_\text{min} = -\infty \), indicating an assisted search guaranteed to fail.

Equation (8) can be written as the difference between two positive numbers:

\[
I_+ = \log_2 \frac{1}{p} - \log_2 \frac{1}{q}.
\]

We call the first term the *endogenous information*, \( I_\Omega \), given in Eq. (1). Endogenous information represents the fundamental difficulty of the search in the absence of any external information. The endogenous information therefore bounds the active information in Eq. (8).

\[-\infty \leq I_+ \leq I_\Omega.\]

The second term in Eq. (9) is the *exogenous information*.

\[
I_\Omega := -\log_2 q.
\]

The endogenous information is the (nonuniform) measure on \( \Omega \) that represents the difficulty inherent in blind search that the assisted search overcomes. Thus, for instance, an assisted search that is no better at locating a target than blind search entails zero active information.

Like other log measurements (e.g., dB), the active information in Eq. (8) is measured with respect to a reference point. Here endogenous information based on blind search serves as the reference point. Active information can also be measured with respect to other reference points. We therefore define the active information more generally as

\[
I_+(\phi|\psi) := \log_2 \frac{\phi(T)}{\psi(T)} = \log_2 \phi(T) - \log_2 \psi(T)
\]

for \( \phi \) and \( \psi \) arbitrary probability measures over the compact metric space \( \Omega \) (with metric \( D \)), and \( T \) an arbitrary Borel set of \( \Omega \) such that \( \psi(T) > 0 \).

### 3.2. Horizontal No Free Lunch

The following No Free Lunch theorem (NFLT) underscores the parity of average search performance.

**Theorem 1: Horizontal No Free Lunch.** Let \( \phi \) and \( \psi \) be arbitrary probability measures and let \( \bar{T} = \{T_i\}_{i=1}^N \) be an exhaustive partition of \( \Omega \) all of whose partition elements have positive probability with respect to \( \psi \). Define active entropy as the average active information that \( \phi \) contributes to \( \psi \) with respect to the partition \( \bar{T} \) as

\[
H_+^\bar{T}(\phi|\psi) := \sum_{i=1}^N \psi(T_i) I_+^\bar{T}(\phi|\psi) = \sum_{i=1}^N \psi(T_i) \log_2 \frac{\phi(T_i)}{\psi(T_i)}
\]

Then \( H_+(\phi|\psi) \leq 0 \) with equality if and only if \( \phi(T_i) = \psi(T_i) \) for \( 1 \leq i \leq N \).

**Proof:** The expression in Eq. (11) is immediately recognized as the negative of the Kullback-Leibler distance [19]. Since the Kullback-Leibler distance is always nonnegative, the expression in Eq. (11) does not exceed zero. Zero is achieved if for each \( i \), \( \phi(T_i) = \psi(T_i) \).

Because the active entropy is strictly negative, any uninformed assisted search \( \phi \) will on average perform worse than the baseline search. Moreover, the degree to which it performs worse will track the degree to which the assisted search singles out and confers disproportionately high probability on only a few targets in the partition. This suggests that success of an assisted search depends on its attending to a few select targets at the expense of neglecting most of the remaining targets.

**Corollary: Horizontal No Free Lunch – No Information Known.** Given any probability measure on \( \Omega \), the active entropy for any partition with respect to a uniform probability baseline will be nonpositive.

**Remarks:** If no information about a search exists, so that the underlying measure is uniform, then, on average, any other assumed measure will result in negative active information, thereby rendering the search performance worse than random search.
Proof: Using $\psi = U$ and Eq. (6), the expression Eq. (11) in Theorem 3.2 becomes
\[ H^T_k (\phi | U) = \sum_{i=1}^{N} U(T_i) I^T_i (\phi | U). \]...... (12)
From Theorem 3.2, the expression in Eq. (12) is nonpositive.

4. The Search for a Good Search

What is the source of active information in a search? Typically, programmers with knowledge about the search (e.g., domain expertise) introduce it. But what if they lack such knowledge? Since active information is indispensable for the success of the search, they will then need to S4S. In this case, a good search is one that generates the active information necessary for success. In this section, we prove a vertical no free lunch theorem. It establishes that, under general conditions, the difficulty of the S4S as measured against an endogenous information baseline, increases exponentially with respect to the minimum active information needed for the original search.

4.1. The Probabilistic Hierarchy

Our first task is to extend the definition of active information in Eq. (10) to the probabilistic hierarchy that sits on top of $\Omega$. Doing so will allow us to characterize information-theoretically “the search for a good search,” “the search for the search for a good search,” etc.

Given $\Omega$, define $M(\Omega)$ as the collection of all probability measures defined on the Borel sets of $\Omega$. We show in Appendix A that $M(\Omega)$ is itself a compact metric space with Borel sets, and thus admits the set of all probability measures on it, which we denote by $M^2(\Omega) = M(M(\Omega))$ and which is again a compact metric space, and so on.

Accordingly, given that $M^0(\Omega) := \Omega$, $M^1(\Omega) = M(\Omega)$, in general $M^k(\Omega) = M(M^{k-1}(\Omega))$ is a compact metric space. $M^k(\Omega)$ is the $k$th order probability space on $\Omega$. This is the probabilistic hierarchy. For notational simplicity, we will use the abbreviation $M^k(\Omega) = M^k$.

To extend the definition of active information in Eq. (10) up the probabilistic hierarchy, we need to show how higher-order probabilities are canonically transformed into lower-order probabilities. Denote an arbitrary probability measure on $M^k$ by $\Theta_{k+1}$ (which is in $M^{k+1}$). By vector-valued integration [20–22]
\[ \Theta_k = \int_{M^k} \mu d\Theta_{k+1}(\mu). \]...... (13)
Active information in $\Omega$ can thus be interpreted as being propagated from active information down the probabilistic hierarchy. Because of the compactness of $\Omega$ and $M(\Omega)$, the existence and uniqueness of Eq. (13) is assured. Note that such integrals exist provided that all continuous linear functionals applied to them (which, in this case, amounts to integrating with respect to all bounded continuous real-valued functions on $M^k$) equals integrating over the continuous linear functions applied inside the integral. Linear functionals thereby reduce vector-valued integration to ordinary integration.

1. Propagation of Active Information from $M(\Omega)$ to $\Omega$. Fig. 1 illustrates the simple problem of searching for one of 16 possible squares in a $4 \times 4$ grid (= $\Omega$). The square in question is the target marked $T$. Here $p = \frac{1}{16}$ and the endogenous information is $I_\Omega = 4$ bits.

Figure 2 illustrates how active information for searching $M(\Omega)$ propagates down the probabilistic hierarchy to active information for searching $\Omega$. Consider the following two probability distributions in $M(\Omega)$:

- $A$ is the uniform distribution over the rightmost four squares in the search space.
- $B$ is the uniform distribution over the bottom twelve squares in the search space.

Suppose we know that one of these distributions characterizes the search for $T$ in $\Omega$, but we know nothing else about the search for $T$. In that case, Bernoulli’s PrOIR would have the search for $T$ characterized by a probability measure $\Theta_2 (\in M^2(\Omega))$ that assigns probability $\frac{1}{4}$ to both $A$ and $B$. By Eq. (13), $\Theta_2$ induces the following probability of success on $\Omega$:
\[ q = Pr[T|A] Pr[A] + Pr[T|B] Pr[B] \]
\[ = \frac{1}{4} \times \frac{1}{12} + \frac{1}{2} \times \frac{1}{2} = \frac{3}{8}. \]
The corresponding exogenous information is therefore $I_s = - \log_2 \frac{3}{8} = 2.59$ bits and the active information is $I_+ = 4.00 - 2.59 = 1.41$ bits. Thus, the higher-order probability measure $\Theta_2$ propagated approximately $1\frac{1}{2}$ bits of active information down the probabilistic hierarchy.
2. **Negative Active Information from \( M(\Omega) \).** If incorrect information assists a search, the active information can be negative. Consider, again, Fig. 1 and assume, as in the previous example, that one of the distributions \( A \) and \( B \) in \( M(\Omega) \) characterizes the search for \( T \), though we have no way of preferring one to the other. This time, as shown in Fig. 3, assume that the target \( T \), instead of being in the lower right corner, is in the lower left corner. The probability of success is then

\[
q = \Pr[T|A] \Pr[A] + \Pr[T|B] \Pr[B]
\]

which corresponds to exogenous information \( I_S = 4.59 \) and to active information \( I_+ = 4.00 - 4.59 = -0.59 \) bits.

3. **Propagation of Active Information from \( M^{2}(\Omega) \).** Fig. 4 illustrates the propagation of active information from \( M^2(\Omega) \) down the probabilistic hierarchy. One of the distributions in \( M^2(\Omega) \) is the Bernoulli distribution \( \Theta_2 \) that assigns \( \frac{7}{8} \) probability to the distribution \( A \) and \( \frac{1}{8} \) to \( B \). Over \( \hat{\Omega} \), the distribution \( \hat{A} \) is uniform on the right four squares of \( \Omega \) and \( B \) is uniform on the bottom eight squares. The probability of success is then

\[
q = \Pr[T|A] \Pr[A] + \Pr[T|C] \Pr[C]
\]

\[
= \frac{1}{4} \times \frac{7}{8} + \frac{1}{8} \times \frac{1}{8} = 0.0156.
\]

Therefore, \( I_S = -\log \frac{15}{64} = 2.09 \) bits corresponding to an active information of \( I_+ = 4.00 - 2.09 = 1.91 \) bits.

### 4.2. Conservation of Uniformity

In the absence of active information, uniform probability measures propagate down the probabilistic hierarchy to lower-level uniform probability measures. In other words, higher-order uniform probability measures never induce anything other than uniform probability measures at lower levels in the probabilistic hierarchy.

**Theorem 2: Conservation of Uniformity.** Let \( U_1 := U \) be the uniform probability on a compact metric space \( \Omega \) with metric \( D \) [23], and let \( U_2 \) be the uniform probability on the compact metric space \( M \) with canonical (Kantorovich-Wasserstein) metric [24]. Then

\[
U_1 = \int_{\Omega} \mu dU_2(\mu), \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (14)
\]

and, more generally,

\[
U_k = \int_{\Omega^k} \mu dU_{k+1}(\mu), \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (15)
\]

where \( U_{k+1} \) is the uniform distribution on \( \Omega^k \).

The proof is in Appendix B.

**Remarks:** Because of the compactness of \( \Omega \) and \( \Omega^1 = \Omega \), the integral in Eq. (14) exists and is uniquely determined. This theorem shows that averaging the probability measures of \( \Omega \) with respect to the uniform probability \( U_2 \) (\( \in \Omega^2 \)) is just the uniform probability \( U_1 \) (\( \in \Omega^1 \)). Thus, if we think of each \( \mu \) under the integral in Eq. (14) as a probability measure representing an assisted search, this theorem says that the average performance of all assisted searches is equivalent to uniform random sampling.

### 4.3. The Displacement Principle: The Vertical No Free Lunch Theorem

The displacement principle states that the search for a good search is at least as difficult as a given search. We prove that in the search for a good search, the endogenous information for the higher-level search grows exponentially with respect to the active information needed to successfully search for the original target. Since endogenous information gauges the inherent difficulty of a search, this shows that the difficulty of the search for a good search grows exponentially with the difficulty of the original search.

More precisely, the original search seeks a target \( T = T_1 \) in \( \Omega = \Omega_1 = \Omega^1 \). In the search for the search, we have a target \( T_2 \) of searches that equal or exceed a given performance level among the set of measures \( \Omega_2 = \Omega^2 \). (We will henceforth interchangeably use the notation \( \Omega_{k+1} \) and \( \Omega^k \).)

Consider, then, a minimally acceptable active information \( \hat{I}_+ \). Over the set of all measures \( \Omega^1 \), the target set of searches is then

\[
T_2 = \{ \varphi \in \Omega_2 \, | \, I_+(\varphi; U) \geq \hat{I}_+ \}. \quad \ldots \quad \ldots \quad (16)
\]

**Theorem 3: The Strict Vertical No Free Lunch Theorem.** Let \( \hat{I}_k = \log(q/p) \) be the minimally acceptable active information of a search such that

\[
p \ll \hat{q} \ll 1 \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (17)
\]

and

\[
p = \frac{1}{K} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (18)
\]

where we have adopted the notation \( K := |\Omega| \). Given that \( M^1 = \Omega_2 \) and that information is measured in nats (as op-

---

1. The subscripts here refer to the level of the search. In Section 2.1, the subscripts refer to multiple queries.
2. A breve as is used on \( \bar{I}_S \) denotes a lower bound while a cap, as is used on \( \bar{q} \), denotes an upper bound.
For a probability, \( \pi \), the endogenous information of a search for a search that achieves at least an active information \( I_+ \) is exponential with respect to \( I_+ \):

\[
I_{\Omega_2} \simeq e^{I_+}.
\]  

(Theorem 4.3)

The proof is supplied in Appendix C.

**Theorem 4: The General Vertical No Free Lunch**

Let

\[
p \ll \hat{q} < 1
\]

and

\[
pK \geq 1.
\]

Then the difficulty of the search for the search as measured by the endogenous information, \( I_{\Omega_2} \), is bounded by

\[
I_{\Omega_2} \geq \hat{I}_{\Omega_2}
\]

where

\[
\hat{I}_{\Omega_2} = \frac{I_0 - \log(K)}{2} - I_+ + KD(p||\hat{q}).
\]  

(23)

Here \( D(p||\hat{q}) \) is the Kullback-Leibler distance:

\[
D(p||\hat{q}) = (1 - p) \ln \left( \frac{1 - p}{1 - \hat{q}} \right) + p \ln \left( \frac{p}{\hat{q}} \right).
\]  

(24)

The proof of Theorem 4.3 is given in Appendix D. Although the strict vertical NFLT in Eq. (18) and the general vertical NFLT in Eq. (21) make different assumptions, the following result shows that Theorem 4.3 is a special case of Theorem 4.3.

**Theorem 5: Strict Case Subsumed in General Case.**

If Eqs. (17) and (18) are in force, then Eq. (23) becomes \( \hat{I}_{\Omega_2} \rightarrow e^{I_+} \), which is consistent with Eq. (19).

The proof is given in Appendix E.

---

5. Conclusion

The Horizontal NFLT illustrates the law of conservatism of information by revealing that unsubstantiated arbitrary assumptions about a search will, on average, result in a search with less than average performance as measured by the search’s active information. This results from the average active information, e.g., the active entropy, being always negative.

---

References:


---

Note: For a probability, \( \pi \), information in bits is given by \( b = -\log_2 \pi \) and, in nats, is \( \eta = -\ln \pi \) where \( \ln \) denotes the natural logarithm [19]. One nat equals \( \ln 2 = 0.693 \) bits.
Polish spaces that can be metrized with a complete metric are a complete separable metric space. Separable topological structures on \( M \) make the Kantorovich-Wasserstein metric \( D_{\Omega} \) makes \( M \) weak topology on \( M \), the probabilistic hierarchy, which is defined inductively as

\[
\frac{\min}{\inf} \left\{ \sum_{i=1}^{n} \delta_{x_{i}} : x_{i} \in \Omega \right\}.
\]

To see how structures on \( M \) extend the probabilistic hierarchy, begin with the metric \( D \) on \( \Omega \). Because \( D \) makes \( \Omega \) a compact metric space, \( D \) a fortiori makes \( \Omega \) a complete separable metric space. Separable topological spaces that can be metrized with a complete metric are known as Polish spaces ([25], ch. 8).

\( M(\Omega) \) is itself a separable metric space in the Kantorovich-Wasserstein metric \( D_{1} \) which induces the weak topology on \( M(\Omega) \). For Borel probability measures \( \mu \) and \( \nu \) on \( \Omega \), this metric is defined as

\[
D_{1}(\mu, \nu) = \inf \left\{ \int D(x, y) \xi(dx, dy) : \xi \in P_{2}(\mu, \nu) \right\}
\]

where \( P_{2}(\mu, \nu) \) is the collection of all Borel probability measures on \( \Omega \times \Omega \) with marginal distributions \( \mu \) on the first factor and \( \nu \) on the second. In the second equation here, \( f \) ranges over all continuous real-valued functions on \( \Omega \) for which the Lipschitz seminorm does not exceed one. The Lipschitz seminorm is defined as follows:

\[
\|f\|_{L} = \sup \left\{ \left| \int f(x) \mu(dx) - \int f(x) \nu(dx) \right| : \|f\|_{L} \leq 1 \right\}
\]

In the first equation, \( P_{2}(\mu, \nu) \) is the collection of all Borel probability measures on \( \Omega \times \Omega \) with marginal distributions \( \mu \) on the first factor and \( \nu \) on the second. In the second equation here, \( f \) ranges over all continuous real-valued functions on \( \Omega \) for which the Lipschitz seminorm does not exceed one. The Lipschitz seminorm is defined as follows:

\[
\|f\|_{L} = \sup \left\{ \left| \int f(x) \mu(dx) - \int f(x) \nu(dx) \right| : \|f\|_{L} \leq 1 \right\}
\]

Both the infimum and the supremum in these equations define metrics. The first is the Wasserstein metric, the second the Kantorovich metric. Though the two expressions appear different, the quantities they represent are known to be identical [26].

The Kantorovich-Wasserstein metric \( D_{1} \) is the canonical extension to \( M(\Omega) \) of the metric \( D \) on \( \Omega \). It extends the metric structure of \( \Omega \) as faithfully as possible to \( M(\Omega) \). For instance, if \( \delta_{x_{i}} \) and \( \delta_{y_{j}} \) are point masses in \( M(\Omega) \), then \( D_{1}(\delta_{x_{i}}, \delta_{y_{j}}) = D(x_{i}, y_{j}) \). It follows that the canonical embedding of \( \Omega \) into \( M(\Omega) \), i.e., \( x \mapsto \delta_{x} \), is in fact an isometry.

To see that \( D_{1} \) canonically extends the metric structure of \( \Omega \) to \( M(\Omega) \), consider the following reformulation of this metric. Let

\[
M_{m}(\Omega) = \left\{ \frac{1}{n} \sum_{i=1}^{n} \delta_{x_{i}} : x_{i} \in \Omega \right\}
\]

where \( n \) ranges over all positive integers. It is readily seen that \( M_{m}(\Omega) \) is dense in \( M(\Omega) \) in the weak topology. Note that the \( x_{i} \)s are not required to be distinct, implying that \( M_{m}(\Omega) \) consists of all convex linear combinations of point masses with rational weights; note also that such combinations, when restricted to a countable dense subset of \( \Omega \), form a countable dense subset of \( M(\Omega) \) in the weak topology, showing that \( M(\Omega) \) is itself separable in the weak topology.

For any measures \( \mu \) and \( \nu \) in \( M_{m}(\Omega) \), it is possible to find a positive integer \( n \) such that

\[
\mu = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_{i}}
\]

and

\[
\nu = \frac{1}{n} \sum_{i=1}^{n} \delta_{y_{i}}.
\]

Next, define

\[
D_{1}^{perm}(\frac{1}{n} \sum_{i=1}^{n} \delta_{x_{i}}, \frac{1}{n} \sum_{i=1}^{n} \delta_{y_{i}}) = \min \left\{ \frac{1}{n} \sum_{i=1}^{n} D(x_{i}, y_{i}) : \sigma \in S_{n} \right\}
\]

where \( S_{n} \) is the symmetric group on the numbers 1 to \( n \). \( D_{1}^{perm} \) looks for the best way to match up point masses for any pair of measures in \( M_{m}(\Omega) \) vis-a-vis the metric \( D \).
It is straightforward to show that $D_{1}^{perm}$ is well-defined and constitutes a metric on $M_{n}^{1}(\Omega)$. The only point in need of proof here is whether for arbitrary measures $\frac{1}{n}\sum_{1 \leq i \leq n} \delta_{x_{i}}$ and $\frac{1}{n}\sum_{1 \leq i \leq n} \delta_{y_{i}}$, and for any measures

$$\frac{1}{mn} \sum_{1 \leq i \leq mn} \delta_{x_{i}} = \frac{1}{n} \sum_{1 \leq i \leq n} \delta_{x_{i}},$$

and

$$\frac{1}{mn} \sum_{1 \leq i \leq mn} \delta_{y_{i}} = \frac{1}{n} \sum_{1 \leq i \leq n} \delta_{y_{i}},$$

we have

$$\min \left\{ \frac{1}{n} \sum_{1 \leq i \leq n} D(x_{i},y_{\sigma i}) : \sigma \in S_{n} \right\} = \min \left\{ \frac{1}{mn} \sum_{1 \leq i \leq mn} D(z_{i},w_{\rho i}) : \rho \in S_{mn} \right\}.$$  

The proof follows from Hall’s marriage lemma [27]. Given this equality, it follows that $D_{1}^{perm} = D_{1}$ on $M_{n}^{1}(\Omega)$ and, because $M_{n}^{1}(\Omega)$ is dense in $M(\Omega)$, that $D_{1}^{perm}$ extends uniquely to $D_{1}$ on all of $M(\Omega)$.

Thus, given that $D$ metrizes $\Omega$, $D_{1}$ is the canonical metric that metrizes $M(\Omega)$. And, just as $D$ induces a compact topology on $\Omega$, so does $D_{1}$ induce a compact topology on $M(\Omega)$. This last result is a direct consequence of $D_{1}$ inducing the weak topology on $M(\Omega)$ and of Prohorov’s theorem, which ensures that $M(\Omega)$ is compact in the weak topology provided that $\Omega$ is compact (28: 59). Given that $D_{1}$ makes $M(\Omega)$ a compact metric space, the next question is whether this metric induces a uniform probability $U_{2}$ on $M(\Omega)$ (such a $U_{2}$ would reside in $M^{2}(\Omega)$). $U_{2} = U$ is the original uniform probability on $\Omega$. As it is, $M(\Omega)$ is uniformizable with respect to $D_{1}$ and therefore has such a uniform probability. To see this, note that

$$U_{2}^{e} = \frac{1}{n} \sum_{1 \leq i \leq n} \delta_{x_{i}}$$

denotes a finitely supported uniform probability that is based on an $\varepsilon$-lattice $\{x_{1},x_{2},\ldots,x_{n}\} \subset \Omega$, then $U_{2}^{e}$ approximates $U$ to within $\varepsilon$, i.e., $D_{1}(U_{2}^{e},U) \leq \varepsilon$. Given that $\binom{2n-1}{n}$ is the number of ways of filling $n$ slots with $n$ identical items ([29], ch. 1), it follows that for

$$n^{*} = \binom{2n-1}{n} = \frac{(2n-1)!}{n!(n-1)!},$$

$$\{\mu_{1},\mu_{2},\ldots,\mu_{n}\} \subset M(\Omega)$$

is an $\varepsilon$-lattice as the $\mu_{k}$s run through all finitely supported probability measures $\frac{1}{n} \sum_{1 \leq i \leq n} \delta_{w_{i}}$ where the $w_{i}$s are drawn from $\{x_{1},x_{2},\ldots,x_{n}\}$ allowing repetitions.

It follows that as $U_{2}^{e}$ converges to $U$ in the weak topology on $M(\Omega)$, the sample distribution

$$U_{2}^{e} = \frac{1}{n} \sum_{1 \leq k \leq n^{*}} \delta_{\mu_{k}}$$

corversges to the uniform probability $U_{2}$ in the weak topology on $M^{2}(\Omega)$. Moreover, for $D_{2}$, which is the Kantorovich-Wasserstein metric on $M^{2}(\Omega)$ ($M^{2}(\Omega)$ is likewise a compact metric space with respect to $D_{2}$), $D_{2}(U_{2},U_{1}) \leq \varepsilon$. The details for proving these claims derive from elementary combinatorics and from unpacking the definition of uniform probability [23].

In summary, given the compact metric space $\Omega$ with metric $D$ and uniform probability $U$ induced by $D$, both $D$ and $U$ extend canonically up the probabilistic hierarchy: $D := D_{0}$ on $M^{0}(\Omega) := \Omega$, $D_{1}$ on $M^{1}(\Omega)$, $D_{2}$ on $M^{2}(\Omega)$, etc. Moreover, $U_{1} \in M^{1}(\Omega)$, $U_{2} \in M^{2}(\Omega)$, etc.

**Appendix B. Proof of the Conservation of Uniformity**

We prove Theorem 4.2 for $\Omega$ finite (the infinite case is proved by considering finitely supported uniform probabilities on $\varepsilon$-lattices, and then taking the limit as the $\varepsilon$-mesh of these lattices goes to zero [23]).

**Proof.** Let $\Omega = \{x_{1},x_{2},\ldots,x_{n}\}$. For large $N$, consider all probabilities of the form

$$\mu = \sum_{1 \leq i \leq N} \frac{N_{i}}{N} \delta_{x_{i}}$$

such that the $N_{i}$s are nonnegative integers that sum to $N$. From elementary combinatorics, there are $N^{*} = \frac{(N+n-1)!}{n!(N-1)!}$ distinct probabilities like this [29]. Therefore, define

$$U_{2}[N] := \frac{1}{N^{*}} \sum_{1 \leq k \leq N^{*}} \delta_{\mu_{k}}$$

so that the $\mu_{k}$s run through all such $\mu$. From Appendix A it follows that $U_{2}[N]$ converges in the weak topology to $U_{2}$ as $N \uparrow \infty$.

It’s enough, therefore, to show that

$$\int_{M(\Omega)} \mu dU_{2}[N] = \frac{1}{N^{*}} \sum_{1 \leq k \leq N^{*}} \mu_{k}$$

is the uniform probability on $\Omega$. And for this, it is enough to show that for $x_{1}$ in $\Omega$,

$$\frac{1}{N^{*}} \sum_{1 \leq k \leq N^{*}} \mu_{k}(\{x_{1}\}) = \frac{1}{n}.$$  

For definiteness, let us consider $x_{1}$. We can think of $x_{1}$ as occupied with weights that run through all the multiples of $1/N$ ranging from 0 to $N$. Hence, for fixed integer $M$ ($0 \leq M \leq N$), the contribution of the $\mu_{k}$s with weight $M/N$ at $x_{1}$ is

$$M \cdot \left( \frac{N-M+n-2}{n-2} \right).$$

Note that the term $\binom{N-M+n-2}{n-2}$ is the number of ways of occupying $n-1$ slots with $N-M$ identical items [29]. Accordingly, the total weight that the $\mu_{k}$s assign to $x_{1}$, when normalized by $1/N^{*}$, is

482 Journal of Advanced Computational Intelligence and Intelligent Informatics Vol.14 No.5, 2010
Because this expression is independent of \( x_1 \) and is also the probability of \( x_1 \), it follows that the probability of all \( x_i \) is under \( \sum_{1 \leq k \leq N^*} \mu_k \) is the same. Hence for each \( x_i \) in \( \Omega \),

\[
U(\{x_i\}) = \frac{1}{N^*} \sum_{1 \leq k \leq N^*} \mu_k(\{x_i\}) = \frac{1}{n}.
\]

This is what needed to be proved.

### Appendix C. Proof of the Strict Vertical No Free Lunch Theorem

We offer two derivations of Theorem 4.3.


We start with the finite case. First, we assume

\[
\Omega = \{x_1, x_2, \ldots, x_K\},
\]

\[
T = \{x_1, x_2, \ldots, x_{|T|}\},
\]

where, from Eq. (6), we use \(|T| = pK\). (For the proof of Theorem 4.3, we use Eq. (18).) For a given \( N \), consider all probabilities

\[
\mu = \frac{1}{N^*} \sum_{1 \leq i \leq N^*} \delta_{w_i}
\]

where the \( w_i \)'s are drawn from \( \Omega \) allowing repetitions. By elementary combinatorics there are

\[
N^* = \binom{N + K - 1}{K - 1} = \frac{\Gamma(N + K)}{\Gamma(N + 1)\Gamma(K)}
\]

such probabilities.\(^4\)

Accordingly, define

\[
U_2[N] := \frac{1}{N^*} \sum_{1 \leq i \leq N^*} \delta_{\mu_i}
\]

where the \( \mu_i \)s range over these \( \mu \)s. \( U_2[N] \) then converges weakly to \( U_2 \).

Next, consider the probabilities

\[
\mu = \frac{1}{N^*} \sum_{1 \leq i \leq N^*} \delta_{w_i}
\]

where the \( w_i \)'s are drawn from \( \Omega \), allowing repetitions, but also for which the number of \( w_i \)'s among \( T \) is at least \( N_q \).

If we let \( Q_N \) denote the set of all such finitely supported probabilities that assign probability at least \( \hat{q} \) to \( T \), and if we let

\[
q(N) = N - \lfloor N\hat{q} \rfloor \xrightarrow{N \to \infty} (1 - \hat{q})N,
\]

then by elementary combinatorics \( Q_N \) has the following cardinality.

\[
|Q_N| = \sum_{j=0}^{q(N)} \binom{N - j + pK - 1}{j + (1 - p)K - 1}.
\]

(25)

Because we assume Eq. (18), this gives

\[
|Q_N| = \sum_{j=0}^{q(N)} \binom{j + K - 2}{K - 2}.
\]

This last limit simplifies. Note first that

\[
\binom{N + K - 1}{K - 1} = \frac{N^{K-1} + (\text{lower order terms in } N)}{\Gamma(K)}.
\]

Thus

\[
\binom{N + K - 1}{K - 1} \xrightarrow{N \to \infty} \frac{N^{K-1}}{\Gamma(K)}.
\]

Similarly, note that

\[
\binom{j + K - 2}{K - 2} = \frac{j^{K-2} + (\text{lower order terms in } j)}{\Gamma(K + 1)}.
\]

Since

\[
\sum_{j=0}^{M} j^n = \frac{M^{n+1}}{n+1} + (\text{lower order terms in } M),
\]

we can write

\[
\sum_{j=0}^{q(n)} \binom{j + K - 2}{K - 2} \xrightarrow{N \to \infty} \sum_{j=0}^{(1 - \hat{q})N} \frac{j^{K-2}}{\Gamma(K + 1)}.
\]

(29)

Substituting Eqs. (17) and (18) into Eq. (26) gives

\[
p_2 \xrightarrow{N \to \infty} (1 - \hat{q})^{K-1}.
\]

The endogenous information for the search for the search

\[
I_{\Omega_2} = -\ln p_2
\]

is then

\[
I_{\Omega_2} = -(K - 1)\ln(1 - \hat{q}).
\]

(32)

\(\Gamma(\cdot)\) is the gamma function. For \( M \) a nonnegative integer, \( \Gamma(M + 1) = M! \). We use \( \Gamma(\cdot) \) rather than the factorial because we do not distinguish between integer and real arguments in our treatment. More rigorously, the beta function \((30)\),

\[
\beta(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}
\]

takes on the role of the binomial coefficient for noninteger arguments.
Fig. 5. A three dimensional simplex in \( \{\omega_1, \omega_2, \omega_3\} \). The numerical values of \( a_1, a_2 \) and \( a_3 \) are one.

Assuming Eq. (17) requires that \( \ln(1 - \hat{q}) \rightarrow -\hat{q} \) and, using Eq. (18)

\[
K - 1 \xrightarrow{p \to 0} \frac{1}{p}.
\]

Therefore Eq. (32) becomes

\[
I_{\Omega_2} \simeq \frac{\hat{q}}{p} = e^{\hat{q}}.
\]

C.2. Geometric Approach

Define the set of all possible probability density functions on \( \Omega \) as

\[
\text{pdf}(\Omega) = \left\{ f(n) \mid 1 \leq n \leq K, \sum_{n=1}^{K} f(n) = 1 \right\}.
\]

The set pdf(\( \Omega \)) lies on a simplex. To illustrate, the simplex equilateral triangle \( \{a_1, a_2, a_3\} \) is shown in Fig. 5 for \( K = 3 \) in the coordinate system \( \{\omega_1, \omega_2, \omega_3\} \). We choose a point on the simplex at random assuming a uniform distribution. Then

\[
\Pr[\hat{q} \geq q] = \Pr[f(0) \geq q].
\]

This is equivalent in Fig. 5 to choosing a point on the triangle \( \{a_1, b_1, b_2\} \). The ratio the area of this smaller triangle to the area of the simplex triangle is the probability a successful pdf will be chosen. The two triangles are congruent. In higher dimensions, the success area will be congruent with the simplex. The ratio scales with the dimension, so

\[
p_2 = \Pr[\hat{q} \geq q] = (1 - \hat{q})^{K-1}. \quad \cdots \cdots \quad (33)
\]

This is equivalent to Eq. (30), from which the rest of the derivation follows.

Appendix D. Proof of the General Vertical No Free Lunch Theorem

Proof of Theorem 4.3.

The preamble to this proof is identical to the development in Appendix C.1 up to Eq. (25), from which point we continue. For \( pK \geq 1 \), it follows that

\[
p_2 := U_2(T_2) = \lim_{N \to \infty} \frac{|Q_N|}{N^*}
\]

\[
= \lim_{N \to \infty} \frac{q(N(K-j+pK-1))}{pK} \frac{\sum_{j=0}^{N-K} \frac{pK}{j+1} \frac{1}{p(1-p)^{K-1}}}{N^{j+1}}.
\]

\[
\cdots \cdots \quad (34)
\]

Similarly, note that

\[
\frac{N-j+pK-1}{pK-1} = \frac{(N-j)p^{K-1} + (\text{lower order terms in } N-j)}{\Gamma(pK)}
\]

and that

\[
\frac{j+(1-p)K-1}{(1-p)K-1} = \frac{j^{(1-p)K-1} + (\text{lower order terms in } j)}{\Gamma((1-p)K)}
\]

so that

\[
\frac{N-j+pK-1}{pK-1} \xrightarrow{N \to \infty} \frac{(N-j)p^{K-1}}{\Gamma(pK)}
\]

and

\[
\frac{j+(1-p)K-1}{(1-p)K-1} \xrightarrow{N \to \infty} \frac{j^{(1-p)K-1}}{\Gamma((1-p)K)}.
\]

Using this and the denominator simplification in Eq. (27), we find that Eq. (34) becomes

\[
\frac{N-j+pK-1}{pK-1} \xrightarrow{N \to \infty} \frac{(N-j)p^{K-1}}{\Gamma(pK)}
\]

and

\[
\frac{j+(1-p)K-1}{(1-p)K-1} \xrightarrow{N \to \infty} \frac{j^{(1-p)K-1}}{\Gamma((1-p)K)}.
\]

This last limit becomes the cumulative beta distribution [30]:

\[
p_2 = \left( \frac{K}{pK} \right) \int_0^{1-\hat{q}} t^{(1-p)K-1} (1-t)^{pK-1} dt. \quad (35)
\]

We now apply Stirling’s exact formula to calculate the factor in front of the integral in Eq. (35). According to Stirling’s exact formula, for every positive integer \( n \),

\[
\sqrt{2\pi n^{n+1/2}} e^{-n} < n! < \sqrt{2\pi n^{n+1/2}} e^{-n+1/(12n)},
\]

which implies that there is a function \( \epsilon(n) \) satisfying \( 0 < \epsilon(n) < 1 \) such that [31]

\[
n! = \sqrt{2\pi n^{n+1/2}} e^{-n+\epsilon(n)/(12n)},
\]

Since \( \Gamma(n+1) = n! \), it now follows for all \( K \) that
\[
\left( \frac{K}{pK} \right) = \frac{\Gamma(K)}{\Gamma(pK)\Gamma((1-p)K)}
\]
\[
= \frac{\Gamma(K)}{\Gamma((1-p)K)\Gamma(pK+1)} (1-p)pK^2
\]
\[
= p(1-p)K
\]
\[
\times \frac{\sqrt{2\pi} K^{1/2} e^{-K+\varepsilon(K)/(12K)}}{\sqrt{2\pi}((1-p)K)^{(1-p)K+1/2}}
\]
\[
\times \frac{e^{(1-p)K-\varepsilon((1-p)K)/(12(1-p)K)}}{\sqrt{2\pi}((1-p)K)^{(1-p)K+1/2}}
\]
\[
= \frac{p(1-p)K}{2\pi} \left( (1-p)(1-p)p^p \right)^{-K}
\]
\[
\times \exp \left[ \frac{\varepsilon(K)}{12K} - \frac{\varepsilon((1-p)K)}{12pK} - \frac{\varepsilon((1-p)K)}{12(1-p)K} \right]
\]
\[
\leq \frac{p(1-p)K}{2\pi} e^{\frac{1}{2}K} \left( (1-p)(1-p)p^p \right)^{-K}
\]
\[
< \sqrt{p(1-p)K} \left( (1-p)(1-p)p^p \right)^{-K}.
\]

Moreover, the integral in Eq. (35) can be bounded for large K.
\[
\int_0^{1-q} t^{(1-p)K-1}(1-t)^{pK-1} dt
\]
\[
\leq (1-\tilde{q})(1-\tilde{q})^{(1-p)K-1}\tilde{q}pK-1
\]
\[
= (1-\tilde{q})^{(1-p)K}\tilde{q}pK-1
\]

It now follows that for large K obeying Eq. (36), we have
\[
P_2 = \left( \frac{K}{pK} \right) \int_0^{1-q} t^{(1-p)K-1}(1-t)^{pK-1} dt .
\]

\[
< \sqrt{p(1-p)K} \left( (1-p)(1-p)p^p \right)^{-K}
\]
\[
\times (1-\tilde{q})^{(1-p)K}\tilde{q}pK-1
\]
\[
= \sqrt{\frac{p(1-p)K}{\tilde{q}^2}} \left( (1-p)(1-p)p^p \right)^{-K}
\]
\[
\times ((1-\tilde{q})(1-p)p^p)^K
\]
\[
= \sqrt{\frac{p(1-p)K}{\tilde{q}^2}} \left[ \left( \frac{1-\tilde{q}}{1-p} \right)^{1-p} \left( \frac{\tilde{q}}{\tilde{q}} \right) \right]^K
\]

Applying the endogenous information in Eq. (31) to the inequality in Eq. (38) gives Eq. (23) and the proof is complete.

Appendix E. Proof that the Strict Vertical NFL is a Special Case of the General

Proof of Theorem 4.3.
When Eq. (17) is true, we will first show that
\[
\frac{D(p||q)}{p} \xrightarrow{p<\tilde{q}} e^+.
\]
Since \( \ln(1-t) \xrightarrow{t \to 0} -t \), Eq. (24) becomes
\[
\frac{D(p||q)}{p} \xrightarrow{p<\tilde{q}} (1-p)(q-p) - pL_+.
\]
Since \( 1-p \xrightarrow{p \to q} 1 \) and \( q-p \xrightarrow{p \to q} q \), the \( pL_+ \) term is dwarfed and we have
\[
\frac{D(p||q)}{p} \xrightarrow{p<\tilde{q}} q.
\]
and
\[
\frac{D(p||q)}{p} \xrightarrow{p<\tilde{q}} \frac{\tilde{q}}{p} = e^+,
\]
which is the result of the strict vertical NFLT in Eq. (19).
Dembski, W. A. and Marks II, R. J.

Name: William A. Dembski

Affiliation: Senior Fellow, Center for Science and Culture, Discovery Institute

Address: 208 Columbia Street, Seattle, WA 98104, USA

Brief Biographical History:
1996- Senior Fellow with Discovery Institute
2000- Associate Research Professor in the conceptual foundations of science at Baylor University
2007- Senior Research Scientist with the Evolutionary Informatics Lab

Main Works:

Membership in Academic Societies:
- American Mathematical Society (AMS)
- Senior Member, Institute of Electrical and Electronics Engineers (IEEE)

Name: Robert J. Marks II

Affiliation: Department of Electrical & Computer Engineering, Baylor University

Address: One Bear Place, Waco, TX 76798-7356, USA

Brief Biographical History:
1977-2003 Department of Electrical Engineering, University of Washington, Seattle, WA
2003- Distinguished Professor of Electrical & Computer Engineering, Baylor University, Waco, Texas

Main Works:

Membership in Academic Societies:
- Fellow, Optical Society of America (OSA)
- Fellow, Institute of Electrical and Electronics Engineers (IEEE)