Uniform distribution is a model assumption

Olle Häggström*

October 2007

Abstract

In a recent paper [10], William Dembski's use of the No Free Lunch (NFL) Theorems for arguing against the feasibility of Darwinian evolution was shown to be flawed. In a response coauthored with Robert Marks [5], Dembski disagrees. The present paper surveys this exchange, and then expands further on the arguments in [10] for why NFL cannot shed light on evolutionary biology. This leads to some general considerations about the role of uniform distributions in scientific modelling.

1 Introduction

Public understanding of science is arguably more important today than ever before. Global warming and genetic engineering are just two of the pressing issues that, on one hand, need to be dealt with in a democratic framework and therefore discussed publically, and, on the other hand, cannot be addressed without a reasonable level of scientific literacy.

Pseudoscience – including the kind of postmodernist thinking that invites us to pick and choose as we like between contradictory descriptions of the world – is a serious obstacle to public understanding of science, and it is my opinion that scientists need to take a firm stand against it. The present paper is meant as a modest contribution in this direction, and directed against certain pseudoscientific claims made in the neocreationist movement known as Intelligent Design.[1]

More specifically, I am concerned with William Dembski's *No Free Lunch* (2002), and in particular with his (ab)use of the so-called No Free Lunch (NFL) Theorems from optimization theory to argue that evolution by natural selection does not work (other parts of his book have been thoroughly debunked by, among others, Elsberry and Shallit [8], Olofsson [14], Perakh [17] and Wein [23]). In fact, I refuted his alleged NFL application already in an earlier paper [10]. However, rather than backing down from his claims, he states in a recent manuscript coauthored with Robert Marks [5] that far from refuting his arguments, my paper "merely restates them". This claim is unsubstantiated, as any competent reader of [10] and [5] will easily find[2] (and as will

---

*Dept of Mathematics, Chalmers University of Technology, `http://www.math.chalmers.se/~olleh/`. Research supported by the Swedish Research Council and by the Göran Gustafsson Foundation for Research in the Natural Sciences and Medicine.

[1] For a critical survey of Intelligent Design, see the recent collection [3].

[2] One such reader is Mark Perakh [18], who writes that "such categorical asseverations, unfortunately for Marks and Dembski, are not supported by any substantive arguments. They are just naked declarations. I see no reason to accept those assertions, but leave the detailed response to Häggström, who certainly is capable of defending his thesis". Thank you, Mark! Here comes.

be clear from later sections). I nevertheless take the response from Dembski and Marks as an excuse to elaborate a bit further on this issue, as there are additional pedagogical points to be made.

Going into heads-on debate with a pseudoscientist is not entirely without risk, however, as we may play straight into a gambit aptly described by Dennett [7]:

> "Smith's work in geology supports my argument that the earth is flat," you say, misrepresenting Smith's work. When Smith responds with a denunciation of your misuse of her work, you respond, saying something like: "See what a controversy we have here? Professor Smith and I are locked in a titanic scientific debate. We should teach the controversy in the classrooms."

I am, obviously, willing to take the risk, as I find the alternative of leaving the claims in [5] unanswered even less appealing. What I will not do, however, is to make the mistake of being overly diplomatic about pseudoscience and thereby rendering it an air of respectability.

The rest of this paper is organized as follows. In Section 2 I will recall the optimization theory that Dembski relies on, followed in Section 3 by his alleged NFL application, my refutation of it, and the Dembski–Marks response. Dembski's position, though not entirely easy to make sense of, appears to involve a dogmatic faith in uniform distributions, and the role of these is therefore discussed in Section 4. Finally, in Section 5, I offer some advice to Dembski on how he could distill and clarify his argument; such clarification is of course something that any serious scientist will want to do, although in Dembski's case it will come at the cost of making the flaws in his reasoning evident also to the lay reader.

2 Optimization and NFL

A natural setup for discrete optimization theory is the following. Let V be a finite set, let S be a finite set of real numbers, and let $f : V \to S$ be some function. Our task is to find some $v \in V$ such that $f(v)$ is large; in the purest of optimization problems we need to find a $v \in V$ that maximizes $f(v)$ over all of V, while here we will be more interested in the problem of finding a v such that $f(v)$ exceeds some given threshold T.

What is a good algorithm for achieving this? The "brute force" approach is to go through all of V systematically, compute $f(v)$ for each $v \in V$, and keep track of successive maxima and for which v they are achieved. Typically, however, the cardinality $|V|$ of V is astronomical (or worse), making brute force impossible in practice, so more clever algorithms are needed. A very common way to proceed is to devise local search algorithms; see, e.g., [1].

As a prototypical example, consider the Travelling Salesman Problem. Given a collection of $n = 100$ (say) cities $c_1, \dots, c_n$ and an $(n \times n)$-matrix of distances between them, the task is to find a permutation π of $(c_1, \dots c_n)$ such that the length $f(\pi)$ of the corresponding journey through them is as small as possible. With 100! permutations to deal with, brute force is obviously out of the question. Local search in the space of permutations is a better idea, but requires a suitable definition of "local". A natural choice is define two permutations to be neighbors if one can be obtained from the other by reversal of some segment in the permutation (this corresponds to rewiring two of the edges in the tour). One basic local search algorithm is to start with an arbitrary permutation π, compute f for π and all of its $n(n-1)/2 = 4950$ neighbors, replace

π by the f-minimizer among these, and repeat. For obvious reasons, this algorithm is called the hill climber. Often (but far from always) hill climbers and related local search algorithms tend to work reasonably well.

The following result of Wolpert and Macready [25] is intended to warn against the unrealistic idea of an all-purpose optimization algorithm that works well on all problems.

Theorem 2.1 (The No Free Lunch Theorem) *Given* V *and* S*, assume that* $f : V \to S$ *is generated at random according to the Pure Noise model. Then no algorithm performs better than blind search.*

To understand the statement of the theorem, some terms need to be explained:

- Blind search means that at each step of the algorithm, an element $v \in V$ is chosen at random according to uniform distribution among those that have not previously been chosen.
- The Pure Noise model means that for each $v \in V$, independently of all others, the value of $f(v)$ is chosen according to uniform distribution on S.
- For our purposes here, we do not need to formulate the full extent what "performs better" means, but in particular the statement of the theorem implies that for any n and for any T, no algorithm improves upon blind search as regards the probability of finding a value that exceeds T amongst the n first $f(v)$'s computed.

Formulated in this way, the result is quite obvious: Since the $f(v)$'s are independent and identically distributed, at no stage of the algorithm can it make any difference (on average) which new element v we choose to compute $f(v)$ for. So no algorithm can perform better than any others in terms of the the probability of finding a value exceeding T amongst the n first $f(v)$'s computed.

Each time I have spoken about my paper [10] to a mathematical audience, sooner or later some member of the audience has asked "Why in the world has such a trivial observation been elevated to the status of a Theorem?".[3] The question is a good one. I cannot offer a complete answer, but believe that the hype was made possible partly by the somewhat less transparent way in which Wolpert and Macready chose to formulate the Pure Noise model. It is well known and easy to show (see [10]) that Pure Noise is the same as saying that f is chosen at random according to uniform distribution on the set V^S of all possible functions from V to S. This is closer to the Wolpert–Macready way of phrasing the result, although they avoid probabilistic language altogether, and state instead that *averaged over all possible* $f : V \to S$*, no algorithm does better than any other.* With such a formulation, the triviality of the statement becomes less obvious (to a non-probabilist), and furthermore the presence of a model assumption is less manifest. In fact, it seems that most commentators of Wolpert and Macready [25] fail to see that there is any model assumption involved, and read it as a statement about general f.

---

[3] In fact, worse than just being given such status, the result has received much publicity and is described as a major breakthrough not only by Wolpert and Macready themselves and by Dembski, but also by others such as Ho and Pepyne [11] who suggest, astonishingly, that the NFL Theorem is to optimization what Gödel's first incompleteness theorem is to mathematics.

3 No free lunch in biology?

Dembski's application of the NFL Theorem to evolutionary biology is based on viewing the Darwinian process of mutation, reproduction and natural selection as a kind of optimization algorithm. Some readers might feel inclined to dismiss this kind of approach due to the contrast between the clean and digital discreteness of computer algorithms on one hand, and the "juiciness" of biology on the other. I believe such a quick dismissal to be mistaken. If proper scientific standards, here meaning in particular a critical view on the relation between mathematical model and reality and an appropriate amount of modesty in interpreting results, are in force, then biology is likely to have a lot to learn from algorithmics. For instance, Dennett employs the algorithmic perspective very consistently in his wonderfully illuminating *Darwin's Dangerous Idea* [6].

For concreteness, and in the language of the previous section, we may take V to be the set of all DNA sequences of length up to 10^{10}, i.e.,

$$V = \bigcup_{i=1}^{10^{10}} \{A, C, G, T\}^i , \qquad (1)$$

and let $S \subset \mathbf{R}$ be a collection of levels measuring degrees of reproductive fitness (in some given environment). The function $f : V \to S$ then describes what biologists call the fitness landscape, and $f(v)$ describes how fit an organism with genome $v \in V$ is. Darwinian evolution may be thought of as a kind of local search algorithm in this fitness landscape, while V is equipped with a natural neighborhood structure such as declaring $u, v \in V$ to be neighbors if one can be obtained from the other by insertion, deletion or altering of a single nucleotide pair. (This tends to ignore things like macromutations and recombination, though the algorithmic model of evolution can of course be refined so as to incorporate such phenomena.)

Blind search in the biological fitness landscape will manifestly not work. An oft-quoted analogy is the tornado in the junkyard: the probability that a tornado sweeping through a junkyard will randomly assemble a full Boeing 747 from the junk is so small that we can safely dismiss the idea that such an event will take place anywhere in the known universe during the next few billion years. Likewise, no complete human being (or any other creature of similar complexity) will ever[4] emerge suddenly out of primordial soup.

This sets the stage for Dembski's [4] "big idea": The NFL Theorem says that no algorithm performs better or worse than any other, so in particular Darwinian evolution is just as bad as blind search. And since the latter is a hopeless means for producing higher life, then so is Darwinian evolution. (*Ergo*, we must be the work of an intelligent designer.)

In [10] I point out the fundamental flaw in Dembski's argument: the failure to address the issue of whether the biological fitness landscape satisfies the critical assumption needed to invoke NFL, namely the Pure Noise model.[5] As I wrote in [10], if the Pure Noise model were an accurate description of the real fitness landscape, then

---

[4] Just to be on the safe side, by "ever" I here mean within the next 10^{100} years or so.

[5] My paper was triggered by the observation that earlier Dembski critics seemed unaware of (the significance of) the Pure Noise assumption, and that instead some confused ideas about what was wrong with Dembski's idea were circulating. In his review of *No Free Lunch*, Orr [15] points out that the actual fitness landscape is not constant over time, claiming that this renders NFL inapplicable. But the Wolpert–Macready paper [25] does contain an analogue of Theorem 2.1 for time-dependent functions (equally trivial as Theorem 2.1), and Dembski [4] in fact addresses this point. In a later paper [16],

> changing a single nucleotide somewhere along the DNA [of a very fit creature such as you or me, would be] just as bad as as putting together a new genome from scratch and completely at random, something that will [...] with overwhelming probability produce not just a slightly less fit creature, but no creature at all. If the true fitness landscape had this property, then, given the human mutation rate, none of us would be around.

This observation alone provides such overwhelming evidence against the true fitness landscape looking anything like what Pure Noise would produce, that a more formal empirical and statistical study of this issue is entirely superfluous.[6] The true fitness landscape clearly exhibits a huge degree of what I in [10] call *clustering*, namely the tendency for large values of the fitness to occur near each other in V far more often than predicted by Pure Noise.

Perhaps surprisingly, Dembski and Marks [5] do not contest this central point in [10], so one has to assume that they accept it. Thus, they accept the conclusion that the assumptions of Theorem 2.1 are not in force, not even remotely so, for the true fitness landscape. Why, then, do they not accept my conclusion that NFL is irrelevant to evolutionary biology?

In fact, there is a way to resolve this apparently contradictory position, and this, for all I understand, is what they do, however obscurely, in their talk in [5] about "active information". Namely, a result of the form "if A, then B" can equivalently be restated as "if $\neg B$, then $\neg A$". Applied to Theorem 2.1, the alternative statement of the result becomes the following.

> If an algorithm is found that works better than blind search, then Pure Noise does not hold. (2)

The application to evolutionary biology can then be succinctly stated by saying that

> since the Darwinian algorithm works so well (incomparably better than blind search), the biological fitness landscape cannot be Pure Noise. (3)

This seems to be basically what Dembski and Marks [5] have in mind (though I doubt whether we will ever hear Dembski state the first seven words of (3) explicitly), and

---

Orr reverts to claiming that NFL does not apply to coevolution of two ore more species living in the same ecosystem (a claim that presumably originates from Wolpert [24], and has been further reproduced by Rasmuson [19] and others). Again, this is a red herring. A variant of NFL (again equally trivial) for coevolution is easily obtained, with Pure Noise translated into uniform distribution on $\tilde{V}^S$, where $\tilde{V} = V \times \{0,1\}^V$, and $(v,b) \in \tilde{V}$ represents a creature with genome $v \in V$ living in an environment containing precisely those creatures whose corresponding digit in the binary string b is 1.

Similarly confused is Tellgren's [21] claim that since, for any v, $f(v)$ is uniform on S, each v stands a reasonable chance of having a good fitness, so that in fact blind search works fairly well under the Pure Noise assumption. But Pure Noise has no such implications. For instance, S might consist of 10^{1000} elements, out of which only 10^{100} correspond to creatures that are actually capable of surviving and reproducing, while the remaining elements are just an absurdly fine-grained discrimination of various levels of being dead meat or coding for no organism at all. With such an S, each $v \in V$ has (under Pure Noise) only a probability of 10^{-900} of having any useful fitness at all. (If such a tilted choice of S seems unnatural, just note (as in [10]) that the proof of Theorem 2.1 does not in fact use the assumption that each $f(v)$ is uniformly distributed, but only that the different $f(v)$'s are independent and identically distributed. This allows for the perhaps more natural choice of a much smaller S while still insisting on as small a probability of getting reasonable fitness as we wish.)

[6] Even though ruling out Pure Noise is itself too easy to be an worthwile research topic, quantitative studies in this direction can still be of interest; see, e.g., [9].

they go on to conclude that since the fitness landscape does not satisfy the Pure Noise assumption, it must have been tampered with by an intelligent designer. While I accept the argument in (3), here is where Dembski and I must part company. They deduce, from absence of Pure Noise, the existence of an intelligent designer, but this requires a dogmatic belief in the idea that without the latter, everything must be uniformly distributed.

In fact, this line of reasoning is hinted at already in *No Free Lunch* [4], and I addressed it briefly in [10], stating only that a deviation from Pure Noise in the direction of clustering hardly requires mysterious explanations, a much more sensible approach being to view it as a manifestation of the wide-spread phenomenon "like causes tend to have like consequences". And of course, biochemistry and genetics provide overwhelming support for this view in the special case of similar genomes leading (often, at least) to similar organisms.

What I did not do, however, in [10], was to discuss further Dembski's idea of uniform distribution as a universal default assumption. In the next section, I will make up for this.

4 On the role of uniform distributions

What, then, do Dembski and Marks [5] have to offer in favor of their idea that in a natural world (i.e., one not cooked up by an intelligent designer), uniform distribution is what we expect to see all around us? Not much. All they do is quote Jacob Bernoulli [2] as saying that "in the absence of any prior knowledge, we must assume that the events [in Ω] ... have equal probability". They could equally well have quoted the following passage of Pierre Simon de Laplace [13]:

> The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability, which is thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all the cases possible.

Bernoulli (1654–1705) and Laplace (1749–1827) were undoubtedly great minds, but they lived in times when probability theory was still in a very primitive stage, and the wealth of tools available to scientists today in devising realistic probability models was simply not around. To be sure, uniform distribution still plays a non-negligible role in probability theory, but for a scientist today to invoke it without reflection – or with a mere reference to some 18:th century authorities – is bad scientific practice.

As I emphasize in the title of this paper, uniform distribution is a model assumption among many alternatives, and it is not always an appropriate choice. It is often argued in Bayesian statistics that when we set out to specify a prior distribution on some space Ω, then uniform distribution is a desirable choice, so as to filter out all prejudice about some $\omega \in \Omega$ being preferable to others. It is however well known (even on the level of undergraduate textbooks) that when the space Ω is *infinite*, then there are several problems. One is that the uniform distribution property fails to be closed under nonlinear transformations. For instance, suppose all we know about a square is that its side-length is between 0 or 1, or equivalently that its area is between 0 and 1. Then we can set up the prior distribution so that the side-length is uniform on $[0,1]$, or so that the area is uniform on $[0,1]$, but we cannot accomplish both things simultaneously in the same model. Another problem is that in many cases, such as $\Omega = \mathbf{R}$, uniform distribution is impossible to define unless we allow the total probability mass to be ∞ rather than 1. Such so-called improper priors can be dealt with mathematically, but much of the philosophical underpinnings of Bayesian statistics is lost.

Perhaps less widely understood is the fact that uniform priors are still highly problematic when Ω is *finite but very large*, as I will argue here.

When Ω has a small number of elements, then uniform distribution on Ω tends to work just fine, an ideal situation being something like the following. Suppose we have a coin which we know is of one of three kinds: either unbiased (heads-probability $p = \frac{1}{2}$) or biased with $p = \frac{1}{4}$ or $p = \frac{3}{4}$. By means of repeated tosses of the coin, we wish to figure out which of the three kinds it is. A natural approach here is to define a uniform prior on $\Omega = \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$, and to update this distribution after each toss using Bayes' formula. This works very well, as the updated (i.e., posterior) distribution converges quickly to a point mass at the true value of p. (For instance, a calculation shows that we can be more than 99.99% sure that, after only 1000 tosses, the true value is assigned a posterior probability of at least 0.9999.)

The situation can be very different when Ω is very large. Consider for instance a situation where we have a black-and-white but very fine-grained image of $n \times n$ pixels with $n = 10^6$. The space Ω of possible images has cardinality $2^{10^{12}}$. This is an astoundingly large number, although note that compared to the space S^V considered by Dembski (where already V as defined in (1) is far beyond astronomical), Ω is pathetically small. Let us imagine a situation where we can only observe parts of the image, and wish to draw inference about the full image. We also wish to avoid prejudice about what the true $\omega \in \Omega$ is. What, then, could be less prejudiced than imposing a uniform prior distribution on Ω?

Well, imagine that we have observed three quarters of the image – let's say the first, second, and third quadrants – and every single pixel in those has turned out to be black. What can we conclude about the fourth quadrant? With the uniform prior distribution, each pixel takes value black or white with probability $\frac{1}{2}$ each, independently of all others. So the total blackness of the first three quadrants does not reveal any information about the fourth. We therefore expect about half of the pixels of the fourth quadrant to be white, and half to be black, and moreover that these are so evenly spread out over the quadrant that no fluctuations from the average grey-level are visible to the naked eye, so the quadrant looks uniformly grey.

So here's what the uniform distribution on Ω means: We expect the entire image to be grey, and we believe so strongly in this that no information about what parts of the image look like can make us back off from our (almost) absolute conviction that all parts that we haven't yet looked at are grey. Rather than achieving our ambitions about lack of prejudice, the uniform distribution turns out to be extremely dogmatic in favor of entirely grey images.

The hopes for uniform distributions to serve us as unprejudiced or somehow canonical priors therefore fails. Attempts have been made, based on Kolmogorov complexity and related concepts, to find prior distributions that do not exhibit the dogmatism of the uniform distribution (see, e.g., Rissanen [20]), but to find such a scheme that is universal and works well in practice appears to be a truly difficult project.

5 Some advice for Dembski

I believe myself to have found, at the end of Section 3, the core of Dembski's reasoning regarding NFL and evolutionary biology (though it is of course possible that in making this claim, I overestimate my ability as a Dembski reader). If I am right about this, then substantial simplifications of his argument can be made, and I will end this paper by outlining these. If Dembski aspires to be a real scientist, then his reaction should be to gratefully embrace any chance to clarify his arguments. We'll see.

First of all, while I accept the soundness of the arguments (2) and (3), they do constitute a terribly roundabout way of obtaining the conclusion that the biological fitness landscape looks nothing like Pure Noise. For that conclusion, it suffices to note, as I do in Section 3 and earlier in [10], that under the Pure Noise model, the birth of a healthy baby would (since, due to mutation and sexual recombination, every baby comes with a genome that is not identical to any of its ancestors' or anyone else's) be no more likely than the tornado in the junkyard producing a Boeing 747. Thus, invoking NFL in order to deduce that the fitness landscape is not Pure Noise serves only to impress readers with fancy-looking mathematical formulas.

Second, nothing Dembski says in favor of uniform distributions, and in particular of Pure Noise, seems to be specific to the biological fitness landscape. Rather, his position, for the best of my attempts to make sense of it, is that in the absence of an intelligent designer, we should expect to find Pure Noise wherever we look.

Provided that we accept this position[7], then demonstrating an enormous deviation from Pure Noise anywhere in our physical world would be enough for Dembski's purpose of proving the existence of an intelligent designer. His choice of the somewhat abstract concept of biological fitness landscape is, for the purpose of clarity and transparence, a very bad one. Better would be to go for something more concrete. I propose geography.

Let us specify a grid on the surface of our planet Earth, such that neighboring grid points are separated by approximately a kilometer. At each grid point, we make a measurement to determine whether we have land or water at the point. Pure Noise in this setting is simply uniform distribution on the set $\{\text{land, water}\}^V$, where V is the set of grid points, or equivalently that each each point independently of all others is in state "land" or state "water" with probability $\frac{1}{2}$. The model implies that with overwhelming probability, the fraction of points in state "land" will be very close to $\frac{1}{2}$, and the fraction of nearest neighbor pairs in the same state will be close to $\frac{1}{2}$.[8] But if we actually make the measurements, then, obviously we will find a much larger proportion of water, as well as a very strong clustering in the sense that the proportion of neighboring grid points in the same state will be way larger than predicted by Pure Noise. The magnitude of the statistical evidence against Pure Noise thus obtained would be so enormous that any reasonable scientist would immediately agree that Pure Noise can be rejected, and this concludes my variant of Dembski's argument.[9]

---

[7] This is a crucial part, because if Dembski cannot make a convincing case for it (something he has not done), then his whole line of reasoning falls apart.

[8] Suppose instead we make more sophisticated measurements, allowing us to distinguish between the three states "land", "freshwater" or "saltwater". Then Pure Noise predicts the fraction of the Earth covered by water to be $\frac{2}{3}$ rather than $\frac{1}{2}$. This kind of contradiction is analogous to the trouble with defining a uniform prior for squares up to a certain size noted in Section 4, and shows in yet another way why it is unreasonable to expect everything around us to be uniformly distributed.

[9] A geophysicist would, at this point, offer a number of natural explanations for clustering of land that do not require an intelligent designer. This is analogous to the biochemical, genetical and embryological

Suppose, however, that we weren't satisfied by the evidence obtained from the kilometer-spaced grid. Well, then we can define a millimeter-spaced grid instead, and obtain even more overwhelming statistical evidence against Pure Noise. We see from this choice of grid that Dembski's belief in Pure Noise predicts (analogously to the entirely grey image in Section 4) that our planet is covered entirely by mud.

The reader has surely realized where this is leading: the universal Pure Noise assumption implies that there is no structure at all in the universe. The question of why there is such structure, together with the closely entangled questions "Why are the laws of nature what they are?" and the deeply frightening "Why is there something rather nothing?", take us into the deepest realms of metaphysics and cosmology. But this is where we end up, given the general character of Dembski's arguments. I warmly welcome Dembski to pursue these deep and difficult questions.[10] But they have little or nothing to do with the details of evolutionary biology (no more, it appears, than with, say, the Newtonian theory of gravity; cf. the elegant satire in [22]). So my final piece of advice to Dembski is that he quits pestering the biologists.

Acknowledgement. I thank Anders Gorm Pedersen for drawing my attention to [9].

References

[1] Aarts, E. and Lenstra, J.K., eds. (1997) *Local Search in Combinatorial Optimization*, Wiley, Chichester.

[2] Bernoulli, J. (1713) *Ars Conjectandi.*

[3] Brockman, J., ed. (2006) *Intelligent Thought: Science versus the Intelligent Design Movement*, Vintage, New York.

[4] Dembski, W.A. (2002) *No Free Lunch: Why Specified Complexity Cannot Be Purchased without Intelligence*, Roman & Littlefield, Lanham, MA.

[5] Dembski, W.A. and Marks, R.J. (2007) Active information in evolutionary search, `http://web.ecs.baylor.edu/faculty/marks/T/Hag2.pdf` downloaded on June 9, 2007. The paper is linked to from Dembski's webpage `http://www.designinference.com/`, where it is described, somewhat optimistically, as "currently under review" (September 12, 2007).[11]

[6] Dennett, D.C. (1995) *Darwin's Dangerous Idea*, Simon & Schuster, New York.

[7] Dennett, D.C. (2005) Show me the science, *New York Times*, August 28.

[8] Elsberry, W. and Shallit, J. (2003) Information theory, evolutionary computation, and Dembski's "complex specified information", `http://www.talkreason.org/articles/eandsdembski.pdf`

[9] Guo, H.H., Choe, J. and Loeb, L.A. (2004) Protein tolerance to random amino acid change, *Proceedings of the National Academy of Sciences* **101**, 9205–9210.

---

explanations of why the biological fitness landscape is clustered.

[10] To be fair, it should be mentioned that he does discuss such cosmological issues in *No Free Lunch.* As a less biased starting point for such investigations, I recommend Kuhn's [12] well-referenced attempt at a systematic survey of possible answers.

[11] **Note added on January 8, 2008:** In a later version of their draft, the title has changed into "The information cost of No Free Lunch"; see `http://cayman.globat.com/~trademarksnet.com/Research/EILab/Publications/Hagg.html`

[10] Häggström, O. (2007) Intelligent design and the NFL theorems, *Biology and Philosophy* **22**, 217–230.

[11] Ho, Y.C. and Pepyne, D.L. (2002) Simple explanation of the no-free-lunch theorem, *Journal of Optimization Theory and its Applications* **115**, 549–570.

[12] Kuhn, R.L. (2007) Why this universe? Toward a taxonomy of possible explanations, *Skeptic* **13**, no. 2, 28–39.

[13] Laplace, P.S. (1814) *Eassai Philosophique sur les Probabilités*, Paris.

[14] Olofsson, P. (2007) Intelligent design and mathematical statistics: a troubled alliance, *Biology and Philosophy*, to appear.

[15] Orr, H.A. (2002) Book review: No Free Lunch, *Boston Review*, summer issue.

[16] Orr, H.A. (2005) Devolution, *The New Yorker*, May 30.

[17] Perakh, M. (2003) *Unintelligent Design*, Prometheus Books, Buffalo, NY.

[18] Perakh, M. (2007) Olle, Tom, and "active information": a brief comment, `http://www.talkreason.org/articles/olle.cfm`

[19] Rasmuson, M. (2006) Deism mot Darwinism, *Folkvett*, no. 1, 6–14.

[20] Rissanen, J. (1983) A universal prior for integers and estimation by minimum description length, *Annals of Statistics* **11**, 416–431.

[21] Tellgren, E. (2006) Free noodle soup, `http://www.talkreason.org/articles/nfl_gavrilets6.pdf`

[22] *The Onion* (2005) Evangelical scientists refute gravity with new 'intelligent falling' theory, August 17.

[23] Wein, R. (2002) Not a free lunch but a box of chocolates, `http://www.talkorigins.org/design/faqs/nfl`

[24] Wolpert, D.H. (2003) William Dembski's treatment of the no free lunch theorems is written in jello, *Mathematical Reviews* MR:1884094 (2003b:00012). Also at `http://www.talkreason.org/articles/jello.cfm`

[25] Wolpert, D.H. and Macready, W.G. (1997) No free lunch theorems for optimization, *IEEE Transactions of Evolutionary Computation* **1**, 67–82.