Free Noodle Soup

E. Tellgren

July 12, 2006

In this note I want to point out a relation between the No-Free-Lunch (NFL) Theorem and some simple models of biological evolution. This point is mathematically trivial, but I believe it nevertheless is of some interest. The NFL theorem characterizes the relative average performance of search algorithms and without being an expert in the field, I believe that its main impact has been to make researchers more cautious in interpreting benchmark results as indicating that one search algorithm is better than another. Via William Dembski [1], one of the most well-known advocates of Intelligent Design (ID), the NFL theorem has also entered into arguments purporting to demonstrate scientifically that some biological features should be attributed to godly intervention (or, possibly, intervention by some other "intelligent agent"). Dembski's use of the NFL theorem has been criticized by several authors (see [5, 9, 10, 11] for a sample of different critical views), but as far as I know it has not been noted that the assumptions used to derive the original NFL theorem, far from enabling the scientific conclusion that evolution needs the helping hand of God, have actually been used in somewhat weaker form in simple models of evolution, models in which the evolutionary performance is not prohibitively bad. Both the NFL theorem and these simple models of evolution will be outlined below.

Expressed in mathematical terms, an optimization problem is the problem of finding the optimum of a function f defined on some search space. Pessimists think of f as a *cost function*, assigning a cost value to each element in the search space, to be minimized. Optimists think of f as an *objective* or *fitness function*, assigning measures of quality, to be maximized.

The NFL Theorem, one of several results about optimization presented by Wolpert and Macready [14], states in its most basic form that if the function f to be optimized is chosen randomly[1] among all possible functions, then all search algorithms have the same statistically expected performance. As emphasized by Häggström [5], to stipulate that

- the function f is chosen randomly

is equivalent to stipulating that

- the values of the function f at different points are chosen independently of each other, and

---

[1] By "chosen randomly" I mean "chosen according to a uniform probability distribution".

- the value of the function f at any given point is chosen randomly from the range of possible fitness values.

A consequence of the statistical independence of the fitness values at any two different locations is that already encountered points in the search space are irrelevant for the decision of which new point to sample next. All search strategies have the same performance as random guessing simply because there is no potential for learning from what is encountered. Indeed, the NFL result for search algorithms [14] is an analogue of earlier results for learning algorithms [12, 13]. Being able to in some sense learn from experience is much more important for learning algorithms than for search algorithms, and in the NFL scenario with a randomly chosen f any search algorithm will in fact achieve a fairly good performance level [2] and, very significantly, the performance is independent of the size of the search space. The result that all search algorithms have the same statistically expected performance has also been obtained under weaker conditions [6, 7]; the absolute performance in those cases depends on the abundances of points with different fitness values.

With this introduction to Wolpert and Macready's NFL result, I want to proceed to point out that the type of uncorrelated fitness landscapes that is produced when the fitness values at different locations are statistically independent is known in the literature as a *rugged fitness landscape*. This class of fitness landscapes is among the simplest to analyze mathematically and has therefore been studied in some detail [8]. Notable results, for a search space corresponding to haploid genotypes with n loci and k alleles at each locus, include that the shortest path to the nearest local maximum is a just few steps (point mutations) long, that the average length of the path actually taken during a pure ascent to a local maximum grows very slowly (logarithmically) in the number of loci (n) and alleles (k). The higher a local maximum lies, the larger its basin of attraction with respect to pure ascent in the fitness landscape, meaning that ascent in a rugged fitness landscape is biased towards eventually ending up in a high local maximum. In later work, Gavrilets [3, 4] has considered the regions of the fitness landscape where the fitness is higher than some threshold c. Mathematically, such a region is written

$$T_f = \{x \in X | f(x) \geq c\}. \tag{1}$$

For Gavrilets the search space X is taken to be of the form $X = \{A_1, ..., A_k\}^{2n}$ representing the possible genotypes of a diploid organism with n loci and k possible alleles at each locus. The threshold c is taken by Gavrilets to be the lowest fitness level for which an organism is still viable, so that T_f is the set of all viable organisms, but the result to be mentioned below is not dependent on any specific choice of c. For example, if the range of fitness values is normalized to go from 0 to 1, one can restrict attention to the top 1% of the fitness range by choosing $c = 0.99$.

What does the set of high-fitness genotypes look like? A few possibilities are that the genotypes are scattered into numerous small disconnected islands, a well-connected web of noodles immersed in a soup of lower-fitness genotypes, a

2

swiss cheese with holes corresponding to lower-fitness genotypes, or just a single big island. The answer is that it depends on the proportion p of fitness values above the threshold c and the dimensionality $2n$ of the genotype space. For a large dimensionality, there is a percolation threshold at

$$p_{\text{thres}} = \frac{1}{2n(k-1)}. \tag{2}$$

When p is larger than $\frac{1}{2}$, the set T_f forms a "swiss cheese" type of structure, with holes in what is otherwise large continuous regions of high-fitness genotypes. More interestingly, when p is small, but above the percolation threshold, the set T_f forms a big well-connected "noodle soup" where the vast majority of all high-fitness genotypes are connected by a web of paths that never dip below the fitness threshold c. Because the number of mutational neighbours increases with increasing dimensionality, so does the number of potential paths and consequently the probability that any two high-fitness genotypes are connected must also increase. Therefore the well-connected "noodle soup" structure becomes easier to obtain the higher the dimensionality of the genotype space. For a genotype space representing 100 loci, the genotypes with fitness values in the top 1% of the range will be well-connected. For a genotype space representing 1000 loci, one can restrict attention to the top 0.1% and still have a well-connected set, and so on.

I want to emphasize the scaling in the size and dimensionality of the search space, because it goes against a common intuition that finding points with high fitness gets more difficult as the search space becomes larger. The reason is simply that a randomly chosen fitness function is a very poor model for features that require many tightly coordinated parts, for fitness landscapes where high-fitness regions are very rare and small, and for other things one might have in mind when forming intuitions that finding points with high fitness is very difficult. A randomly chosen fitness function, as in the NFL scenario, does not at all fit with that. Instead, a very special choice of fitness function is required if a model that corresponds to such intuitions is desired.

To summarize, the implications of the assumption of a randomly chosen fitness function do not just include Wolpert and Macready's NFL result, but also the results

- that the absolute performance of any search for high-fitness genotypes is fairly good and, importantly, independent of the size of the genotype space, and

- the set of high-fitness genotypes is well-connected and the connectedness increases with increasing dimensionality of the genotype space.

More metaphorically, the NFL scenario may deny biological evolution a free lunch, but once the lunch break is over it hands evolution a large free bowl of noodle soup.

References

[1] W. A. Dembski. *No Free Lunch: Why specified complexity cannot be purchased without intelligence.* Rowman & Littlefield Publishers Inc., Lanham, MD, 2002. (*Two reactions to critics are also available at* `http://www.designinference.com/documents/05.02.resp_to_wein.htm` *and a* `http://www.designinference.com/documents/2002.12.Unfettered_Resp_to_Orr.htm`. *Recent texts at* `http://www.designinference.com` *provide a somewhat different argument, which merits a different treatment and will not be dealt with here beyond the observation that as long as uniform probability distributions are unreservedly assumed the conclusions about absolute performance and connectedness are difficult to avoid.*).

[2] T. M. English. Optimization is easy and learning is hard in the typical function. In *Proceedings of the 2000 Congress on Evolutionary Computation CEC00*, pages 924–931, La Jolla Marriott Hotel, La Jolla, California, USA, July 2000. IEEE Press.

[3] S. Gavrilets. Evolution and speciation on holey adaptive landscapes. *Trends in Ecology & Evolution*, 12(8):307–312, August 1997.

[4] S. Gavrilets and J. Gravner. Percolation on the fitness hypercube and the evolution of reproductive isolation. *Journal of Theoretical Biology*, 184(1):51–64, January 1997.

[5] O. Häggström. Intelligent design and the NFL theorems. (manuscript available at `http://www.math.chalmers.se/~olleh/papers.html`), 2006.

[6] C. Igel and M. Toussaint. On classes of functions for which No Free Lunch results hold. *Information Processing Letters*, 86(6):317–321, June 2003.

[7] C. Igel and M. Toussaint. No free lunch theorem for non-uniform distributions of target functions. *Journal of Mathematical Modelling and Algorithms*, 3(4):313–322, December 2004.

[8] C. A. Macken, P. S. Hagan, and A. S. Perelson. Evolutionary walks on rugged landscapes. *SIAM Journal on Applied Mathematics*, 51(3):799–827, June 1991. (*Interestingly, in what is perhaps an indication that the NFL result is obvious in hindsight but not so easy to think of oneself unless one thinks to ask the right question, these authors conclude from the bias of ascent towards high local maxima that this is a better search strategy than random search (p. 810). Alternatively, the tension with the general NFL result might be resolved if these authors determine what is better using criteria that only take into account the end result, not the number of points sampled to get there.*).

[9] H. A. Orr. Book review: No free lunch. *Boston Review*, 2002.

[10] M. Perakh. There is a free lunch after all: William Dembski's wrong answers to irrelevant questions. In M. Young and T. Edis, editors, *Why Intelligent Design Fails*, chapter 11. Rutgers University Press, 2004.

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

[12] D. H. Wolpert. The existence of a priori distinctions between learning algorithms. *Neural Computation*, 8(7):1391–1420, October 1996.

[13] D. H. Wolpert. The lack of a priori distinctions between learning algorithms. *Neural Computation*, 8(7):1341–1390, October 1996.

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