

Bernoulli's *Principle of Insufficient Reason* and Conservation of Information in Computer Search

William A. Dembski
Professor of Philosophy
Southwestern Baptist Theological Seminary
Fort Worth, Texas

Robert J. Marks II
Distinguished Professor
Electrical & Computer Engineering Dept.
Baylor University, Waco, Texas

Abstract—*Conservation of information* (COI) popularized by the *no free lunch theorem* is a great leveler of search algorithms, showing that on average no search outperforms any other. Yet in practice some searches appear to outperform others. In consequence, some have questioned the significance of COI to the performance of search algorithms. An underlying foundation of COI is *Bernoulli's Principle of Insufficient Reason*¹ (PrOIR) which imposes of a uniform distribution on a search space in the absence of all prior knowledge about the search target or the search space structure. The assumption is conserved under mapping. If the probability of finding a target in a search space is p , then the problem of finding the target in any subset of the search space is p . More generally, all some-to-many mappings of a uniform search space result in a new search space where the chance of doing better than p is 50-50. Consequently the chance of doing worse is 50-50. This result can be viewed as a confirming property of COI. To properly assess the significance of the COI for search, one must completely identify the precise sources of information that affect search performance. This discussion leads to resolution of the seeming conflict between COI and the observation that some search algorithms perform well on a large class of problems.

Index Terms—active information, Bernoulli's principle of insufficient reason, Bernoulli's principle of non-sufficient reason, Bertrand's paradox, conservation of generalization performance, conservation of information, endogenous information, equal distribution of ignorance, evolutionary search, no free lunch theorem, principle of indifference, uniform distribution

I. INTRODUCTION

*Conservation of information*² [13], [14], [15], [16], [20], [21], [54], [55], [56], [57], [58], [65], [66], [67] (COI) popularized by the *no free lunch theorem* (NFLT) [18], [38], [45], [68], dictates that, on average, one search technique performs as well as any other. Like Gödel's incompleteness theorem, the NFLT reveals an impossibility, namely, that there exists no "magic bullet" search algorithm [38], [11] that performs better, on average, than any other search algorithm. Gödel's result establishes important limitations for computer science, as revealed in the *Turing halting problem* [10] and *Rice's theorem* [52]. Likewise, COI establishes important limitations for search algorithms, as revealed in the amount of active information that must be infused into a search for it to be successful.

¹Also called *Bernoulli's Principle of Non-sufficient Reason* [44], the *Principle of Indifference* [44], and the *Equal Distribution of Ignorance* [41].

²The term *conservation of information* applied to search limitations, including the NFLT, was apparently coined by English [20]. Schaffer [58] earlier used the phrase *A Conservation Law for Generalization Performance*.

Schaffer [58], for example, notes

"... a learner [without prior knowledge] ... that achieves at least mildly better-than-chance performance ... is like a perpetual motion machine."

Concerning the NFLT, Ho and Pepyne [38] write

"... unless you can make prior assumptions about the ... [problems] you are working on, then no search strategy, no matter how sophisticated, can be expected to perform better than any other."

According to Wolpert and Macready [68], search can be improved only by

"... incorporating problem-specific knowledge into the behavior of the [optimization or search] algorithm."

The "prior assumptions" and "problem specific knowledge" required for "better-than-chance performance" in evolutionary search [24], [27], [49], [70] derives from prior knowledge that, when properly fitted to the search algorithm, favorably guides the search.

An underlying assumption of COI is Bernoulli's *principle of insufficient reason* (PrOIR) [2], [50] which states

"... in the absence of any prior knowledge, we must assume that the events [in a search space] ... have equal probability."

Laplace subscribed to the principle when he wrote uniformity should be assumed [23]

"... [when] we have no reason to believe any particular case should happen in preference to any other"

Bernoulli's PrOIR, offered in 1713, is equivalent to an initial imposition of maximum (information-theoretic) entropy on the search space [10], [50]. The assumption of maximum entropy is useful in optimization and is, for example, foundational in Burg's algorithm [4], also known as the *maximum entropy method* [50]. To generate useful solutions using Burg's algorithm, the unknown portion of the problem is constrained to be at maximum entropy.

Most criticisms of COI are aimed at the uniformity assumption imposed by Bernoulli's PrOIR [17], [32], [64] where the meaning of "absence of any prior knowledge" is often misconstrued. Indeed, once imposed, the uniformity assumption is robust on mappings of the search space. We show this new COI result in Section III.

II. CONSERVATION OF INFORMATION

Like many important insights, reflection and hindsight reveals the obviousness of the COI. After presenting his paper *A Conservation Law for Generalization Performance* [58], Cullen Schaffer noted [59]

“About half of the people in the audience to which my work was directed told me that my result was completely obvious and common knowledge—which is perfectly fair. Of course, the other half argued just as strongly that the result wasn’t true.”

A debate also occurred in regard to Gödel’s theorem which, in its original form [29], [30], was understood only by mathematicians and logicians. Turing’s application of the theorem in the *halting problem* [12] and subsequent contributions and popularization in algorithmic information theory by Chaitin [5], [6] reveal both the generality, obviousness and ubiquity of Gödel’s insight.

To illustrate the obviousness of COI, consider the following illustration [32]. If we enter a room where cards from a well shuffled standard 52 card poker deck are laid randomly face down on a table, our chances of turning over the ace of spades ($A\spadesuit$) in less than five card flips is not dependent on the ordering of the cards turned. Using the result of a flipped card, say the $K\clubsuit$, in any way to determine the location of the next move also does not improve the probability of success. After five card flips, no matter how clever the method used, the probability, p , of choosing the $A\spadesuit$ is the same:

$$p = \frac{\binom{51}{4}}{\binom{52}{5}} = 0.0962 \quad (1)$$

A. Information Measures

The *endogenous information*, I_Ω , of a search [14], [15], [16] expresses the difficulty of a search in terms of forecasting coin flips.

$$I_\Omega = -\log_2 p \text{ bits.} \quad (2)$$

An endogenous information of $I_\Omega = 10$ bits, for example, corresponds to a search with a difficulty equivalent to forecasting the outcome of 10 sequential flips of a fair coin. The search can be a single query or a multiple query sequence. For the example of finding the $A\spadesuit$ in five card flips, the endogenous information for the five query search is, from (1) and (2), $I_\Omega = 3.38$ bits. This endogenous information is invariant to the choice of the search algorithm.

The endogenous information can be visualized as resulting from a search space with $N = 1/p$ equally likely outcomes. The expression in (2) is then the information associated with choosing a single target element from this space assuming a uniform distribution [10], *i.e.* assuming Bernoulli’s PrOIR.

Knowledge about the structure of a search space or the location of a target can improve the chances of success to some $q \geq p$. To illustrate, suppose the cards in the search for the $A\spadesuit$ example are arranged into four quadrants of 13 cards and each quadrant contains randomly arranged cards of the same,

albeit unspecified, suit. This knowledge increases the chance³ of finding the $A\spadesuit$ in five card flips from p to $q = 0.274$. The difficulty of this problem has been reduced to the *exogenous information* $I_S = -\log_2 q$ bits. For the card flipping example, $I_S = 1.87$ bits. The decrease in number of bits of difficulty is the *active information* [14], [15], [16], I_+ , extracted from application of the prior knowledge about the problem.

$$I_+ = I_\Omega - I_S = -\log_2 \left(\frac{p}{q} \right). \quad (3)$$

For the card flipping example, $I_+ = 1.51$ bits.

The active information defined in (3) has intuitively satisfying special conditions. To generate a search with success probability $q = p$ requires no active information. To craft a search with twice the probability of success ($q = 2p$) requires a single bit of active information, etc. For a perfect search ($q = 1$), the required active information is the endogenous information and all the available information has been extracted from the search. If misleading knowledge is used,⁴ the search can be made worse than random and the active information can be negative.

Like other log measurements such as dB, active information is measured with respect to a reference value. In the card flipping example, active information was measured given the assumption the poker deck size was fixed at 52 cards. There was no consideration that the deck of cards was chosen from a bin containing Uno, pinochle, and Old Maid card decks, nor was the number of cards in the deck considered to be a random variable. Active information was, instead, measured from the reference of using an ordinary poker deck of 52 playing cards.

Alternately, suppose the poker cards are not only divided into quadrants as before, but all face cards are identified. The active information can be measured with respect to the quadrant division knowledge alone, or with respect to the original random arrangement of all 52 cards.

Critics of Bernoulli’s PrOIR include Keynes [44], who appeals to Bertrand’s paradox [50], [47] as a counterexample. More recent interpretations fault the definition of *random* in Bertrand’s paradox rather than a failure of Bernoulli’s PrOIR [47], [50]. Bernoulli’s PrOIR does not work when a search space is ill-defined or heuristically uncertain as is typical in the social sciences [41], [42], *e.g.* assigning the primary cause of juvenile delinquency to (a) poverty, (b) peer pressure, or (c)

³This follows from elementary probability. Assume we choose among the quadrants until we uncover a spade. If, in quadrant one, we uncover a \clubsuit , quadrant one is never visited again. When a \spadesuit is uncovered, the remainder of card flips, as needed, are spent in the \spadesuit quadrant. Let \spadesuit_k be the event of first revealing a spade on flip k . Then $\Pr[\spadesuit_k] = 1/4$ for $k = 1, 2, 3, 4$. From the *theorem of total probability*,

$$q = \Pr[A\spadesuit] = \sum_{k=1}^4 \Pr[A\spadesuit | \spadesuit_k] \Pr[\spadesuit_k] = \frac{1}{4} \sum_{k=1}^4 \Pr[A\spadesuit | \spadesuit_k]$$

Since $\Pr[A\spadesuit | \spadesuit_k] = (6 - k)/13$, we obtain $q = 0.274$

⁴In the card flipping example, for example, assume the $A\spadesuit$ is placed in the \clubsuit quadrant and the $A\clubsuit$ is placed with the \spadesuit ’s. If this information is not known and all cards in each quadrant are assumed to be of the same suit, the chance of success in five card flips will fall below p and the active information will be negative.

other causes does not warrant a 1/3, 1/3, 1/3 split and, furthermore, contradicts a separate test between only (a) poverty and (b) other causes. Frequentists argue that with no prior knowledge of the search space, assignment of probabilities is inappropriate and falls outside of the scope of probability theory [41]. Yet Bernoulli's PrOIR is often inescapable. If you have one of a million lottery tickets, the best estimate of the probability of winning is one in a million. The probability calculation in (1) also assumes Bernoulli's PrOIR. With no prior knowledge about a target or the search space structure, Bernoulli's PrOIR is likewise applicable to the finite discrete search spaces in computer search.

III. CONSERVATION OF BERNOULLI'S PROIR

The uniform distribution imposed by Bernoulli's PrOIR in a discrete search space remains robust under transformation or manipulation of the search domain. No arrangement of the search space can add active information if there is no knowledge concerning the search target or the search space structure.

This result seems contrary to an objection by Keynes, a critic of Bernoulli's PrOIR. He offers the following criticism [42], [44]:

“Consider the specific volume of a given substance. Let us suppose that we know the specific volume to lie between 1 and 3, but that we have no information as to whereabouts in this interval its exact value is to be found. The Principle of Indifference [Bernoulli's PrOIR] would allow us to assume that it is as likely to lie between 1 and 2 as between 2 and 3; for there is no reason for supposing that it lies in one interval rather than in the other. The specific density is the reciprocal of the specific volume, so that if the later is v the former is $\frac{1}{v}$. Our *data* remaining as before, we know that the specific density must lie between 1 and $\frac{1}{3}$, and, by the use of the Principle ... as before, that it is as likely to lie between 1 and $\frac{2}{3}$ as between $\frac{2}{3}$ and $\frac{1}{3}$.”

Keynes argues that Bernoulli's PrOIR cannot apply to both the specific density and the specific volume because the 50-50 division results differ in each case. Gädenfors and Sahlin [28] note that many criticisms on Bernoulli's PrOIR, such as Keynes's, focus on cases where the underlying random variable has an infinite number of possibilities. Bertrand [3] was “so much impressed by the contradictions of geometrical probability that he wishes to exclude all examples in which the number of alternatives is *infinite*” [44].

For finite discrete search spaces, however, in the absence of any information about the location of the target, Bernoulli's PrOIR is conserved in all some-to-many mappings (see Figure 1) of the search space in the sense that, if the probability of success in the original search space under uniformity is p , then the mapped search space, no matter how configured, will have a probability of success \hat{p} such that

$$\Pr[\hat{p} \geq p] = \frac{1}{2}. \quad (4)$$

Therefore, independent of p , with no knowledge of about the target or the search space structure, we have a 50-50 chance of doing better in all some-to-many mappings and consequently a 50-50 chance of doing worse. Finite discrete search spaces underlie No Free Lunch Theorems [68].

Equation (4) is proved in Appendix A.

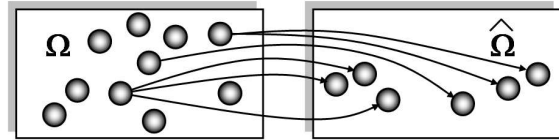


Fig. 1. An example of a some-to-many mapping of a search space Ω to $\hat{\Omega}$.

When considering the example of finding $A\spadesuit$ in a well shuffled deck, the result in (4) is obvious. Choose a subset of cards and, to simulate the equivalent of allowing duplicated cards in the map, assign each subset card a probabilistic weight of importance. With respect to randomly choosing from the original 52 cards, choosing from this weighted subset will not improve your chances of finding the $A\spadesuit$.

A. Identifying a Good Search

In order to craft a search that performs better than uniform randomness, active information is required. Let $q \geq p$ denote an acceptable probability of the new search. How much information is required to create a search-for-a-search with a probability of success of at least q ?

Let Ω_2 be the finite space of all search algorithms on the search space Ω . Let

$$T_2 \subset \Omega_2 \quad (5)$$

be the set of search algorithms that have a probability of success of at least q . Then the endogenous information of the search-for-the-search is

$$I_{\Omega_2} = -\log_2 \left(\frac{|T_2|}{|\Omega_2|} \right). \quad (6)$$

Not surprisingly,

$$I_{\Omega_2} \geq I_+. \quad (7)$$

The search-for-the-search is therefore at least as difficult as the original search active information being sought.⁵ A derivation of (7) is in Appendix B.

IV. INTERPRETING BERNOULLI'S PROIR

In most cases, the controversy over COI can be traced to the interpretation and understanding of Bernoulli's PrOIR, a concept that can be muddled by what may be called *familiarity zones*. Like comfort zones, familiarity zones can become so ingrained that we take for granted things we have no right to take for granted. The “absence of *any* prior knowledge” required for uniformity conceptually parallels the difficulty of

⁵The bound in (7) is conservative. Under loose conditions, the endogenous information of the search-for-the-search can be shown to be $I_{\Omega_2} \simeq e^{I_+}$ when information is measured in nats [14], *i.e.* using a natural logarithm instead of \log_2 .

understanding the *nothing* that physics says existed before the Big Bang. It's common to picture the universe before the Big Bang is a large black void empty space. No. This is a flawed image. Before the Big Bang there was *nothing*. A *large black void empty space* is something. So space must be purged from our visualization. Our next impulse is then, mistakenly, to say, "There was nothing. Then, all of a sudden..." No. That doesn't work either. *All of a sudden* presupposes there was time and modern cosmology says that time in our universe was also created at the Big Bang.⁶ The concept of *nothing* must exclude conditions involving time and space. *Nothing* is conceptually difficult because the idea is so divorced from our experience and familiarity zones.

A similar problem occurs with the phrase "no prior information." Some critics of COI build on the *no free lunch* metaphor and claim that, in some instances, there may be a "free appetizer" [17] or "free leftovers" [9]. They appeal to searches with "links" in the optimization space and smoothness constraints that enable "hill-climbing" optimization [32].⁷ Prior knowledge about the smoothness of a search landscape required for gradient based hill-climbing,⁸ is not only common but is also vital to the success of some search optimizations.⁹ Such procedures, however, are of little use when searching to find a sequence of, say, 7 letters from a 26-letter alphabet to form a word that will pass successfully through a spell checker, or when choosing a sequence of commands from 26 available commands to generate a logic operation such as XNOR [46]. The ability of a search procedure to work better than average on a class of problems is not prohibited by COI.¹⁰ Knowledge about the class in which a problem lies is, indeed, prior knowledge. Such knowledge can come through insight into the search space structure, or it can come from some assessment of the target location or fitness.

Although commonly used evolutionary algorithms such as particle swarm optimization [19] and genetic algorithms [31] perform well on a wide range of problems, there is no discrepancy between the successful experience of practitioners with such versatile algorithms and the COI imposed inability of the search algorithms themselves to create information [7], [20]. The additional information often lies in the experience of the programmer who, with prior knowledge of the problem class, prescribes how the knowledge about the search is to be

⁶For a more thorough high-level explanation, see Hawking [34].

⁷The criticisms of COI can be inappropriately extreme. An example is "Wolpert and Macready [who coined the term NFLT [68]] (1997) contribute their share to the hype [about COI]. And with astonishing lack of perspective, [M.I.T. Ph.D. and Harvard mathematics professor Yu-Chi] Ho and [David] Pepyne (2002) compare the basic NFL theorem to one of the deepest achievements in 20th century mathematics: Gödel's incompleteness theorem." [32]

⁸*E.g.*, steepest ascent and the Widrow-Hoff algorithm [51].

⁹*E.g.*, adaptive filters and the training of layered perceptron neural networks [51].

¹⁰There is a useful analogy in the halting problem [10] where Turing proves there exists no meta-procedure to analyze *any arbitrary* computer program to determine whether the program will stop or not. This does not prohibit, however, meta code from determining whether halting will occur in a *class* of computer programs.

folded into the search algorithm.¹¹ COI takes issue, rather, with claims that one search procedure invariably performs universally better than another, or that remarkable results are due to the search procedure alone [1], [33], [37], [14], [48], [53], [60], [62].

The "no prior knowledge" cited in Bernoulli's PrOIR is all or nothing: we have prior knowledge about the search or we don't. Active information [14], [15], [16], on the other hand, measures the degree to which prior knowledge can contribute to the solution of a search problem. Culberson [11] illustrates the same property in considering the computational requirements of graph coloring dependent on available prior knowledge about the problem. When a graph is known, optimal coloring can be performed in constant time. As the knowledge about the graph decreases, the problem eventually becomes NP-complete and, ultimately, with no knowledge about the graph, the problem requires *blind search* and becomes distinctly "intractable" [43]. Active information allows for the quantification of these different levels of *a priori* knowledge about a search.

A. Co-Evolution

Co-evolutionary optimization uses evolutionary tournament play among agents to evolve winning strategies and has been used impressively to identify winning strategies in both checkers [25] and chess [26]. Although co-evolution has been identified as a potential source for an evolutionary free lunch [69], strict application of Bernoulli's PrOIR suggests otherwise.

To illustrate the applicability of Bernoulli's PrOIR to co-evolution, consider a variation of the game of checkers dubbed "give away" where the winner forces the opponent to capture all of their checkers. The contestant who loses all checkers is declared the winner. The winning strategy in give-away is the diametric opposite of conventional checkers where the winner captures all of the checkers of the opponent. Since the strategies are opposite, a successful strategy for one of the games will miserably fail for the other. Claims for a possible "free lunch" in co-evolution [69] assume prior knowledge of what constitutes a win. Detailed analysis of co-evolution under the constraint of Bernoulli's PrOIR remains an open problem.

B. Biological Implications of COI

Finally, because biological evolution is widely supposed to be Darwinian and thus nonteleological,¹² some critics of COI have argued that it is inapplicable to biology [32], [69]. In particular, they argue that "targets," as humanly imposed artifacts, have no biological analogue. We note, however, that *intrinsic* targets can be mandated by search space structure and are consistent with Cambridge paleobiologist Simon Conway Morris' observation [8] that evolution seems to converge on only a few biological endpoints, *e.g.* the camera eye of humans and squids. He therefore theorizes that if the evolutionary

¹¹Early practitioners of computer search referred to themselves as "penalty function artists" [36].

¹²A point we dispute [13], [16].

process were restarted from the beginning, the life forms of today, including humans, would re-evolve. From the perspective of COI, these limited number of endpoints on which evolution converges constitute intrinsic targets, crafted in part by initial conditions and the environment.

V. THE IMPACT OF THE NFLT

The impact of COI on the field of evolutionary algorithms has been significant. Here are some of its immediate consequences.

- COI has led to the formulation of active information as a measure that needs to be introduced to render an evolutionary search successful [15]. Like an athlete on steroids, many such programs are doctored, intentionally or not, to succeed [14].
- Christensen and Oppacher note the NFLT is “very useful, especially in light of some of the sometimes outrageous claims that had been made of specific optimization algorithms” [7]. Search algorithms do not contribute information to the search, and the COI exposes the inappropriateness of such claims.
- COI shows that claims about one algorithm outperforming another can only be made with regard to benchmarks set by particular targets and particular search structures. Performance attributes and empirical performance comparisons cannot be extrapolated beyond such particulars. There is no all-purpose “magic bullet” search algorithm for all problems [11], [63].
- COI puts to rest the inflated claims for the information-generating power of evolutionary simulations such as Avida [22], [46] and *ev* [60].

VI. CONCLUSIONS

Prior knowledge, when properly prescribed, successfully guides an evolutionary search to a solution by incorporating knowledge about the target and the underlying structure of the search space. That structure determines how the search deviates from the uniformity assumption of the NFLT. References to “geographical structure[s],” “link structure[s],” search space “clustering,” and smooth surfaces conducive to “hill climbing” reinforce rather than refute the quasi-teleological conclusion that the success of evolutionary search depends solely on active information from prior knowledge [32]. This suggests that in biology, as in computing, there is no free lunch [13].

APPENDIX A: PROOF OF EQUATION (4)

Let Ω denote a discrete search space with $|\Omega|$ elements and an unknown target $T \in \Omega$. Under Bernoulli’s ProIR, $p = |T|/|\Omega|$. Consider, then, a some-to-many mapping re-configuration of the search space, $\hat{\Omega} = g(\Omega)$, as illustrated in Figure 1.

Consider, first, the special case of a some-to-one mapping where $\hat{\Omega} \subset \Omega$ and elements in Ω appears in $\hat{\Omega}$ at most once. For a fixed but unknown p , $|\hat{T}|$ is a binomial random variable,

$$\Pr [|\hat{T}| = \hat{k}] = \binom{|\hat{\Omega}|}{\hat{k}} p^{\hat{k}} (1-p)^{|\hat{\Omega}|-\hat{k}}.$$

In the new space, $\hat{\Omega}$, what is the chance that we do better, *i.e.*, that $\hat{p} \geq p$ where $\hat{p} = |\hat{T}|/|\hat{\Omega}|$? We first note that $\Pr [\hat{p} \geq p] = \Pr [|\hat{T}| \geq p|\hat{\Omega}|]$. Since $|\hat{T}|$ is a binomial random variable its median¹³ is $p|\hat{\Omega}|$ [35]. Thus $\Pr [\hat{p} \geq p] = \frac{1}{2}$.

The generalization to cases where elements of Ω can appear more than once in $\hat{\Omega}$ is straightforward. Since the location of the target is unknown, the replication of any element in $\hat{\Omega}$ to make the space larger will result in a new space where the probability of success, \hat{p} , remains the same.

APPENDIX B: PROOF OF EQUATION (7)

Each $\omega \in \Omega_2$ has a measure μ_ω equal to the probability of search success. Then the average probability of success of a search program chosen at random is $p = (1/|\Omega_2|) \sum_{\omega \in \Omega_2} \mu_\omega$. The subset $T_2 \subset \Omega_2$ is the set of search algorithms that have a probability of success of at least q , *i.e.* $T_2 = \{\omega | \mu_\omega \geq q\}$.

Define $\pi = (1/|T_2|) \sum_{\omega \in T_2} \mu_\omega$. From the definition of T_2 it follows that

$$\pi \geq q. \quad (8)$$

Since $\mu_\omega \geq 0$, we have from (5), $\sum_{\omega \in \Omega_2} \mu_\omega \geq \sum_{\omega \in T_2} \mu_\omega$. Therefore $p|\Omega_2| \geq \pi|T_2|$. Since, from (8), $\pi|T_2| \geq q|T_2|$, we have $p|\Omega_2| \geq q|T_2|$. Thus $p/q \geq |T_2|/|\Omega_2|$ from which (7) follows after taking the negative log of both sides.¹⁴

REFERENCES

- [1] L.E. Atlas, R. Cole, Y. Muthusamy, A. Lippman, G. Connor, D.C. Park, M. El-Sharkawi & R.J. Marks II, “A performance comparison of trained multi-layer perceptrons and classification trees”, in **Neural Networks, Theoretical Foundations & Analysis**, C. Lau, editor, pp.284-288, IEEE Press (1992), reprinted from Proceedings of the IEEE, vol.78, pp.1614-1619 (1990).
- [2] Jakob Bernoulli, “Ars Conjectandi (The Art of Conjecturing),” (1713).
- [3] Joseph Bertrand, **Calcul Des Probabilités**, (1896)
- [4] J. P. Burg, “Maximum entropy spectral analysis,” Ph.D. dissertation, Stanford University, Stanford CA, May 1975.
- [5] G. J. Chaitin. **The Limits of Mathematics**, Springer-Verlag, Singapore, 1997.
- [6] G. J. Chaitin. **The Unknowable**, Springer-Verlag, Singapore, 1999.
- [7] S. Christensen and F. Oppacher, “What can we learn from No Free Lunch? A First Attempt to Characterize the Concept of a Searchable,” Proceedings of the Genetic and Evolutionary Computation (2001).
- [8] Simon Conway Morris, **Life’s Solution: Inevitable Humans in a Lonely Universe**, Cambridge University Press (September 8, 2003).
- [9] David W. Corne and Joshua D. Knowles, “No Free Lunch and Free Leftovers Theorems for Multiobjective Optimisation Problems,” in **Evolutionary Multi-Criterion Optimization** (Lecture Notes in Computer Science), Springer Berlin / Heidelberg (2003).
- [10] Thomas M. Cover and Joy A. Thomas, **Elements of Information Theory**, 2nd Edition, Wiley-Interscience, 2006.
- [11] Culberson, J. C. 1998. “On the futility of blind search: an algorithmic view of ‘no free lunch,’” *Evolutionary Computation*, vol. 6, no. 2, pp. 109-127.
- [12] Martin Davis, **The Undecidable, Basic Papers on Undecidable Propositions, Unsolvable Problems And Computable Functions**, Raven Press, New York, 1965.
- [13] William A. Dembski, **No Free Lunch: Why Specified Complexity Cannot Be Purchased without Intelligence**. Rowman & Littlefield Publishers, Inc., 2006.

¹³The median, more rigorously, is either $\lceil p|\hat{\Omega}| \rceil$ or $\lfloor p|\hat{\Omega}| \rfloor$ [35].

¹⁴This proof is a variation of a similar proof given by the authors [16]. The authors express their appreciation to George Montañez for bringing this easily understood version of the proof to our attention.

- [14] W.A. Dembski and R.J. Marks II, "The Search for a Search: Measuring the Information Cost of Higher Level Search," *International Journal of Information Technology and Intelligent Computing*, Vol. 3, No. 4, 2008.
- [15] W.A. Dembski and R.J. Marks II, "Conservation of Information in Search: Measuring the Cost of Success," *IEEE Transactions on Systems, Man and Cybernetics A, Systems and Humans*, September, 2009.
- [16] W.A. Dembski and R.J. Marks II, "Life's Conservation Law: Why Darwinian Evolution Cannot Create Biological Information," in Bruce Gordon and W.A. Dembski, editors, **The Nature of Nature**, (Wilmington, Del.: ISI Books, 2010).
- [17] Stefan Droste, Thomas Jansen, and Ingo Wegener "Perhaps Not a Free Lunch But At Least a Free Appetizer," Technical Report ISSN 1433-3325 September 1998 University of Dortmund, Dept. of Computer Science/XI, 44221 Dortmund, Germany
- [18] Richard O. Duda, Peter E. Hart, and David G. Stork, **Pattern Classification**, Wiley-Interscience; 2 edition (2000).
- [19] Russell C. Eberhart, Yuhui Shi, James Kennedy, **Swarm Intelligence**, (Morgan Kaufmann 2001).
- [20] T.M. English, "Some information theoretic results on evolutionary optimization," *Proceedings of the 1999 Congress on Evolutionary Computation*, 1999, CEC 99. Volume 1, 6-9 July 1999.
- [21] T.M. English, "Evaluation of Evolutionary and Genetic Optimizers: No Free Lunch," in **Evolutionary Programming V: Proceedings of the Fifth Annual Conference on Evolutionary Programming**, L. J. Fogel, P. J. Angeline, and T Bäck, Eds., pp. 163-169. Cambridge, Mass: MIT Press, 1996.
- [22] Winston Ewert, William A. Dembski and R.J. Marks II, "Evolutionary Synthesis of Nand Logic: Dissecting a Digital Organism," *Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics*, October 11-14, 2009, San Antonio, Texas, USA.
- [23] Arne Fisher, Charlotte Dickson, and William Bonyng, **Mathematical Theory Of Probabilities & Its Applications To Frequency Curves & Statistical Methods**, Macmillan (1922)
- [24] David B. Fogel, **Evolutionary Computation: The Fossil Record**, IEEE Press, 1998
- [25] D. B. Fogel, **Blondie24: Playing at the Edge of AI**, Morgan Kaufmann; 1st edition (2001)
- [26] D. B. Fogel, T. J. Hays, S. L. Hahn, and J. Quon, "A self-learning evolutionary chess program," *Proc. IEEE*, vol. 92, pp. 1947-1954, 2004.
- [27] David B. Fogel, **Evolutionary Computation: Toward a New Philosophy of Machine Intelligence**, Third Edition, IEEE Press, 2005
- [28] Peter Gädénfors and Nils-Eric Sahlin **Decision, Probability, and Utility**, Cambridge University Press (1988)
- [29] Kurt Gödel, "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme," *Monatshefte für Mathematik und Physik* 38: 173-98 (1931).
- [30] Kurt Gödel, "Zum intuitionistischen Aussagenkalkül," *Anzeiger Akademie der Wissenschaften Wien* 69: 6566 (1932).
- [31] Tomasz Dominik Gwiazda, *Genetic Algorithms Reference*, (Tomasz Gwiazda, 2006)
- [32] Olle Häggström, *Biology and Philosophy* 22 (2007): 217-230.
- [33] Kuk-Hyun Han and Jong-Hwan Kim, "Quantum-inspired evolutionary algorithm for a class of combinatorial optimization," *IEEE Transactions on Evolutionary Computation*, Volume: 6, Issue 6, pp.580- 593 (2002).
- [34] Stephen W. Hawking, **A Brief History Of Time - From The Big Bang To Black Holes**, Bantam Books (1988).
- [35] K. Hamza, "The smallest uniform upper bound on the distance between the mean and the median of the binomial and Poisson distributions," *Statist. Probab. Lett.* vol.23, pp. 2125 (1995).
- [36] Michael Healy, private communication.
- [37] R. Hinterding, Z. Michalewicz, A.E. Eiben, "Adaptation in evolutionary computation: a survey," *Proceedings of the IEEE International Conference on Evolutionary Computation*, 13-16 Apr 1997 pp. 65-69 (1997).
- [38] Yu-Chi Ho and D.L. Peppyne, "Simple explanation of the No Free Lunch Theorem", *Proceedings of the 40th IEEE Conference on Decision and Control*, Orlando, Florida, 2001.
- [39] Yu-Chi Ho and D.L. Peppyne, "Simple explanation of the no-free-lunch theorem," *J. Optimiz. Theory Appl.* 115: 549570 (2002).
- [40] Yu-Chi Ho, Qian-Chuan Zhao, and D.L. Peppyne, "The No Free Lunch Theorems: Complexity and Security," *IEEE Transactions on Automatic Control*, vol. 48, #5, 2003, pp.783-793.
- [41] David Howie, **Interpreting Probability: Controversies and Developments in the Early Twentieth Century** Cambridge University Press, (2002)
- [42] Edward Kasner & James Roy Newman, **Mathematics and the Imagination**, Dover Publications (2001)
- [43] Neil D. Jones, **Computability and Complexity: From a Programming Perspective**, MIT Press (1997).
- [44] John Maynard Keynes, **A Treatise On Probability**, Macmillan Co., 1921.
- [45] M. Koppen, D.H. Wolpert, W.G. Macready, "Remarks on a recent paper on the 'no free lunch' theorems", *IEEE Transactions on Evolutionary Computation*, June 2001, Volume: 5, Issue: 3, pp. 295 - 296.
- [46] Richard E. Lenski, Charles Ofria, Robert T. Pennock and Christoph Adami, "The evolutionary origin of complex features," *Nature*, vol 423, 139-144 (8 May 2003).
- [47] R.J. Marks II, **Handbook of Fourier Analysis and Its Applications**, Oxford University Press (2009).
<http://HandbookOfFourierAnalysis.com/>
- [48] R. Mendes, J. Kennedy, J. Neves, "The fully informed particle swarm: simpler, maybe better," *IEEE Transactions on Evolutionary Computation*, Volume: 8, Issue: 3, pp.204- 210, (June 2004).
- [49] Marimuthu Palaniswami, Yianni Attikiouzel, Robert J. Marks II, David Fogel and Toshio Fukuda; Editors, **Computational Intelligence: A Dynamic System Perspective**, IEEE Press, 1995.
- [50] A. Papoulis, **Probability, Random Variables, and Stochastic Processes**, 3rd ed. New York: McGraw-Hill, pp. 537-542, 1991.
- [51] Russel D. Reed and R.J. Marks II, **Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks**, MIT Press, Cambridge, MA, 1999.
- [52] H.G. Rice, "Classes of Recursively Enumerable Sets and Their Decision Problems," *Trans. Amer. Math. Soc.* 74, 358-366, 1953.
- [53] T.P. Runarsson and Xin Yao, "Stochastic ranking for constrained evolutionary optimization," *IEEE Transactions on Evolutionary Computation*, Volume: 4, Issue 3, pp. 284-294 (2000).
- [54] Cullen Schaffer, "Overfitting Avoidance as Bias (preliminary version)," *Proceedings of Machine Learning: IJCAI Workshop on Discovery in Databases*, (1991).
- [55] Cullen Schaffer, "Deconstructing the Digit Recognition Problem," in *Machine Learning: Proceedings of the Ninth International Conference (ML92)*, D. Sleeman and P. Edwards Editors, Morgan Kaufmann, 1992
- [56] Cullen Schaffer, "Sparse Data and the Effect of Overfitting Avoidance in Decision Tree Induction," *Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI-92)*, 1992.
- [57] Cullen Schaffer, "Overfitting Avoidance as Bias," *Machine Learning*, Volume 10, Issue 2 (February 1993) Pages: 153 - 178
- [58] Cullen Schaffer, "A conservation law for generalization performance," in *Proc. Eleventh International Conference on Machine Learning*, H. William and W. Cohen. San Francisco: Morgan Kaufmann, 1994, pp.295-265.
- [59] Cullen Schaffer, Private Communication.
- [60] T.D. Schneider, "Evolution of Biological Information," *Nucleic Acids Research*, Vol. 28, No 14, pp.2794-2799 (2000).
- [61] Devinderjit Sivia and John Skilling, **Data Analysis: A Bayesian Tutorial**, Oxford University Press, 2 edition (2006)
- [62] Daniel M. Sunday, "A very fast substring search algorithm," *Communications of the ACM*, Volume 33, Issue 8, pp.132 - 142 (August 1990).
- [63] Alan Turing, "On computable numbers, with an application to the Entscheidungsproblem," *Proceedings of the London Mathematical Society*, Series 2, 42 (1936), pp 230-265 and vol. 43. pp.544-546.
- [64] B. Weinberg and E.G. Talbi, "NFL theorem is unusable on structured classes of problems," *Congress on Evolutionary Computation, CEC2004*. Volume 1, 19-23 June 2004 pp. 220 - 226 Vol.1
- [65] David H. Wolpert, "On overfitting avoidance as bias." Technical Report SFI-TR-92-03-5001, Santa Fe Institute. 1992.
- [66] David H. Wolpert, "On the connection between in-sample testing and generalization error," *Complex Systems* 6: pp.47-94 (1992)
- [67] David H. Wolpert, "Stacked generalization," *Neural Networks* 5:241-259 (1992).
- [68] David H. Wolpert, William G. Macready, "No free lunch theorems for optimization," *IEEE Trans. Evolutionary Computation* 1(1): 67-82 (1997).
- [69] David H. Wolpert, and W.G. Macready, "Coevolutionary Free Lunches," *IEEE Transactions on Evolutionary Computation*, December 2005, Volume 9, Issue 6, pp. 721-735.
- [70] Jacek Zurada, R.J. Marks II and C.J. Robinson; Editors, **Computational Intelligence: Imitating Life**, (IEEE Press, 1994).