OPEN ACCESS

BIO-complexity

Monograph

The Law of Conservation of Information: Search Processes Only Redistribute Existing Information

William A. Dembski[1]*
[1] Discovery Institute, Seattle, WA

Abstract

Conservation of information sparked scientific interest once a recurring pattern was noticed in the evolutionary computing literature. In grappling with the creation of information through evolutionary algorithms, this literature consistently revealed that the information outputted by such algorithms always needed first to be programmed into them. Thus, the primary goal of this literature—to uncover how information could be created *from scratch* or *de novo*—was shown to be misconceived: the information was not created but instead shuffled around or smuggled in, implying that it already existed in some form or other. Information output in these situations therefore always presupposed a counterbalancing input of prior information.

Once this pattern was seen, the next logical step was to quantify the amount of information inputted and outputted, demonstrating a consistent mathematical relation between the two. This led to the proof of a number of theorems about search. In these theorems, a *baseline search* with probability p of success gave way to an *improved search* with probability q of success. Typically p would be very small and close to zero, implying a practically impossible search (like searching for a needle in a haystack). By contrast, q would be much larger and close to one, implying an eminently doable search. The punchline of these theorems was that, as the improved search became itself the subject of a search (a *search for a search*, or S4S), the probability of finding it could not exceed p/q, rendering success of the improved search no more probable than success of the original baseline search, in effect filling one hole by digging another.

Such conservation-of-information theorems, as they came to be called, were search-space specific, adapted to different kinds of search across a range of search spaces. There was a measure-theoretic theorem in which probability measures guided search. There were also function-theoretic and fitness-theoretic theorems where mappings into the search space as well as fitness functions on the search space respectively guided search. The key insight of this paper is that all these conservation-of-information theorems are special cases of a simple probabilistic relation based on elementary probability theory. This paper identifies the underlying rationale that makes all the previous conservation-of-information theorems work. In so doing, it provides a straightforward proof and general formulation of what may rightly be called the Law of Conservation of Information.

Cite As: Dembski WA (2025) The Law of Conservation of Information: Search Processes Only Redistribute Existing Information. BIO-Complexity 2025 (2):1-58. **doi:**10.5048/BIO-C.2025.2.

Editor: Robert J. Marks II

Received: December 13, 2024; **Accepted:** May 28, 2025; **Published:** September 13, 2025

Notes: A *Critique* of this paper, when available, will be assigned **doi:**10.5048/BIO-C.2025.2.c.

*Email: teleocentric@gmail.com

1. THE COST TO INCREASE A PROBABILITY

"Conservation of information" is a term that appears in the physics as well as computer science literature. In physics, it is concerned with physical processes that are in some sense reversible, allowing one physical state to be recovered from another and vice versa. Information is thereby conserved between states when one state is recoverable from the other. In computer science, conservation of information is concerned with search processes for which information output cannot exceed prior information input. Information is thereby conserved when output matches input, unconserved when output falls short of input.

These two uses of the term intersect insofar as physical processes are involved in performing search. Nonetheless, how conservation of information is understood in physics is quite different from how it is understood in computer science. In physics, whenever the word "conservation" is used, whatever is being conserved is supposed to remain strictly constant—no gain, no loss. Conservation of energy in a closed system, for instance, means that the total energy of that system stays strictly constant, though the forms that the energy in the system takes can change (potential, kinetic, chemical, electrical, nuclear, etc.).[1] By contrast, when it comes to search, conservation of information denotes an ideal limit (upper bound) that search processes might achieve but often miss in practice.[2]

In the context of search, conservation of information is best understood as an accounting tool for tracking and capping the informational costs required to raise a smaller probability to a larger one. In ordinary life, we often take steps to raise the probability of some desired outcome. For instance, if you are a student trying to get into an elite college, you want to raise your probability of admission. You or your parents may therefore pay money to a company like Princeton Review to help raise your standardized test scores. You may also invest time and effort to raise your GPA, to take a high number of Advanced Placement classes, and to do extracurriculars—all to improve your probability of acceptance. Getting into an elite school is a crapshoot, but people pay a lot to load the dice in their favor.

Paying to increase the probability of achieving a certain result is quite common. But for most payments that increase probabilities, there is no sense that anything in the payment method is being conserved. It is not as though investing several hundred dollars to improve your standardized test scores corresponds to an equivalent return of that money if your monetary investment helps get you into your desired school. If anything, the return on investment of that money spent on improving your test scores will be amply multiplied if it makes the difference in getting you into a top school, whose graduates will typically earn generous salaries.[3]

In addressing the cost of raising probabilities, conservation of information does not conceive of costs in the ordinary sense as denominated in money, time, or effort. With conservation of information, information itself is treated as a cost. Conservation of information can be understood as an accounting tool, providing a balance sheet in which informational outputs are never able to exceed informational inputs. The costs that conservation of information keeps track of are therefore informational rather than monetary. With conservation of information, to the degree that anything is conserved, it is not money but information.

What, then, is information? Information at its most fundamental is about narrowing possibilities and thus about reducing uncertainty. The more uncertainty is reduced, the greater the information. If I tell you that it is raining today in Seattle, I have given you information. But because it rains there so often, I have not given you much information. The reduction in uncertainty in this case is small. If I tell you that it is raining today in the Sahara desert, I have also given you information. But because it rains there so infrequently, I have given you a lot more information. The reduction in uncertainty in this case is large.[4]

Usually, information theorists measure information in number of bits (a bit being a 0 or 1). But a number of bits is just a disguised form of probability. If we toss a coin 10 times, the probability of the coin landing exactly as it does is $\frac{1}{2}$ times itself 10 times, which comes to a probability of roughly 1 in a thousand. If we toss a coin 100 times, the probability of the coin landing exactly as it does is $\frac{1}{2}$ times itself 100 times, which comes to a probability of roughly 1 in a million trillion trillion. As these probabilities indicate, the reduction of uncertainty in the latter is much more than in the former, and so the latter embodies much more information than the former. If we now represent coin tosses by a sequence of 0s and 1s, or bits (0 for tails, 1 for heads), 10 coin tosses correspond to 10 bits and 100 coin tosses correspond to 100 bits. Provided we now allow fractional bits, any probability can be identified with a number of bits and vice versa. Long story short, information is interpreted as a reduc-

[1] The term "conservation of energy" goes back to 1851 to William Thomson (Lord Kelvin). For a brief history of conservation of energy, see [1].

[2] For a history of the idea of conservation of information in search, see Appendix 1 to this paper. See also [2, ch. 5].

[3] "The research shows that graduates of elite universities tend to earn more than other college graduates. For example, Ivy League graduates earned $161,888 mid-career compared to $101,777 for those from other institutions. Graduates from the top 10 colleges in the United States earn 47% higher early career salaries than those from the ten colleges within the City University New York (CUNY) school system." https://www.hceducationconsulting.com/2023/09/13/unlocking-prosperity-the-higher-earning-potential-of-elite-college-graduates/ (last accessed December 9, 2024).

[4] For a popular account of information as the reduction of uncertainty, see [3, pp. 219–220]:

> Information is uncertainty, surprise, difficulty, and entropy: "Information is closely associated with uncertainty." Uncertainty, in turn, can be measured by counting the number of possible messages. If only one message is possible, there is no uncertainty and thus no information. Some messages may be likelier than others, and information implies surprise. Surprise is a way of talking about probabilities. If the letter following *t* (in English) is *h*, not so much information is conveyed, because the probability of *h* was relatively high. "What is significant is the difficulty in transmitting the message from one point to another." Perhaps this seemed backward, or tautological, like defining mass in terms of the force needed to move an object. But then, mass can be defined that way. Information is entropy. This was the strangest and most powerful notion of all. Entropy—already a difficult and poorly understood concept—is a measure of disorder in thermodynamics, the science of heat and energy.

With the quotes and key terms in this passage, Gleick is quoting [4].

tion of uncertainty and is measured by a probability, or its equivalent in bits.[5]

2. A FIRST EXAMPLE OF CONSERVATION OF INFORMATION

To illustrate how conservation of information works, consider the tossing of a fair die. The probability of rolling a six is $\frac{1}{6}$. But now imagine that you have the power to load the die so that any two distinct numbers are bound to come up, and either of the numbers that could come up will then have the same probability, which is to say $\frac{1}{2}$. Suppose, therefore, that the die has been loaded to come up either a five or a six, each with probability $\frac{1}{2}$. You have now increased your probability of rolling a six. Instead of having a probability of $\frac{1}{6}$ of rolling a six as you did with the fair die, you now have a probability of $\frac{1}{2}$ using the weighted die. That is a significant improvement in the probability of successfully getting a six.

Now, how did you achieve this level of improved probability? If you can load the die one way, you can presumably load it other ways—the geometry and probabilistic properties of the die do not care how the faces are labeled. It is, therefore, safe to assume you could load the die so that any two faces come up, each with probability $\frac{1}{2}$. To increase your probability from $\frac{1}{6}$ to $\frac{1}{2}$ of rolling a six, you could, instead of five-or-six, have loaded the die to come up four-or-six, three-or-six, two-or-six, or one-or-six. Thus, you had five ways to load the die so that with probability $\frac{1}{2}$ you roll a six. But you also had ten other ways of loading the die that would have ensured that you will not roll a six: one-or-two, one-or-three, one-or-four, one-or-five, two-or-three, two-or-four, two-or-five, three-or-four, three-or-five, or four-or-five. With any of these other ways of loading the die, your probability of rolling a six would go down to 0.

Because of the geometric symmetry of the die, any of these ways of loading the die are probabilistically equivalent, so the most reasonable probability distribution to put over these ways of loading the die will make them all equiprobable. In general, conservation of information does not need to presuppose uniform probability distributions like this with equiprobable outcomes. In the present case, however, that distribution is appropriate. Consequently, loading the die on five-or-six is one among five probabilistically equivalent ways of loading the die so that it comes up a six with probability $\frac{1}{2}$. But there are also ten other probabilistically equivalent ways that you could have loaded the die so that it comes up a six with probability 0, which is to say that a six is guaranteed not to be rolled.

It follows that only one-third of the ways of loading the die raise your probability of rolling a six to $\frac{1}{2}$, the other two-thirds lower that probability to 0. Accordingly, if you randomly selected from among these dice-loading methods, you would have a $\frac{1}{3}$ chance of selecting one that boosted the chances of rolling a six to $\frac{1}{2}$. Thus, at a probability "cost" of $\frac{1}{3}$, you were able to load a die that with probability $\frac{1}{2}$ rolls a six. Because of the way these probabilities operate in sequence, they need to be multiplied together to determine the probability of getting a six: You need first to succeed at loading the die in the right way so that a six has a better than zero chance of happening at all. Getting such a properly loaded die has probability $\frac{1}{3}$. And then you will need to roll such a properly loaded die so that it lands a six. Doing so has probability $\frac{1}{2}$.

Multiplying these two probabilities ($\frac{1}{3} \times \frac{1}{2}$) yields a probability of $\frac{1}{6}$. But this is just the original probability of rolling a six with a fair die. It follows that this die-loading scheme has not at all helped to increase the probability of rolling a six. You initially had a probability of $\frac{1}{6}$ of rolling a six. With the die-loading scheme, you have a $\frac{1}{3}$ probability of loading it in a way that with probability $\frac{1}{2}$ rolls a six, making the probability of getting a six likewise to be $\frac{1}{6}$. So, procuring a better weighted die that with higher probability rolls a six has gained you nothing because of the probabilistic cost incurred in procuring it.

How does this example look if we recast it in terms of information conceived as the reduction of uncertainty? In this case, rolling the original unloaded die has a probability of $\frac{1}{6}$ of landing a six. In landing a six, we see a reduction of uncertainty from six parts to one part. In rolling a loaded die that has probability $\frac{1}{2}$ of landing a six, we see instead a reduction of uncertainty from two parts to one part—a considerably smaller reduction in uncertainty and therefore a less information-intensive die roll. Yet, to roll such a loaded die, we need to sample from different ways of loading it, and getting a successful loading of the die (one that will land a six with probability $\frac{1}{2}$) has probability $\frac{1}{3}$. This represents a reduction in uncertainty from three parts to one part, with a corresponding contribution of information. Tying everything together, we see that neither approach to reducing uncertainty outperforms the other if the aim is to roll a six.

In this example, the broad outlines of conservation of information become evident. In situations where conservation of information applies, raising a probability always incurs a probabilistic cost that erases any benefit from raising it. As we just saw, the probability of getting a six using an ordinary die is $\frac{1}{6}$. And yet, the probability of getting a six using a weighted die, once the probability of getting an appropriately weighted die is factored in, is again $\frac{1}{6}$. In this scheme, information is therefore actually conserved, maintaining the original probability. According to conservation of information, such schemes

[5] [5] provides a detailed treatment of the relationship between probability and information. It is a standard reference in the field of information theory.

for raising probabilities at best maintain parity with the originally calculated probability, in this case $\frac{1}{6}$. In fact, such schemes may do worse than maintain parity, leaving information unconserved.

To see how such die-loading schemes can do worse than maintain parity with the original probability and thus leave information unconserved, consider a variation of the previous example. Let ε denote a tiny number greater than 0 (suppose it is less than $\frac{1}{1{,}000}$). Now consider all die-loading schemes in which any two faces can have probability $\frac{1}{2} - \varepsilon$ of being rolled and the other four faces then each have probability $\frac{\varepsilon}{2}$ of being rolled. The probabilities of all the faces then sum to 1, as is easily confirmed. Moreover, there are, as before, fifteen loadings, only this time the two faces with bigger weights have probability just less than $\frac{1}{2}$ (i.e., $\frac{1}{2} - \varepsilon$) and the other four faces with smaller weights have probability just greater than 0 (i.e., $\frac{\varepsilon}{2}$).

In this example, as before, five loadings confer a bigger probability on rolling a six (i.e., $\frac{1}{2} - \varepsilon$), and the other ten loadings confer a smaller probability on rolling a six (i.e., $\frac{\varepsilon}{2}$). Rolling a six directly has, as before, a probability of $\frac{1}{6}$. And similar to before, finding a loading that rolls a six with probability $\frac{1}{2} - \varepsilon$ has probability $\frac{1}{3}$. In this modified example, $\frac{1}{3}$ is the probabilistic cost to make rolling a six more probable than $\frac{1}{6}$. But this time, for the same probabilistic cost, the probability of rolling a six is diminished: rolling a six directly has probability $\frac{1}{6}$, but rolling a six by first finding a loading that with probability $\frac{1}{2} - \varepsilon$ rolls a six now has probability $\left(\frac{1}{2} - \varepsilon\right) \times \frac{1}{3} = \frac{1}{6} - \frac{\varepsilon}{3}$, which is strictly less than $\frac{1}{6}$.

Why is strict conservation of information absent in this example but not in the previous one? In both examples, there are "good loadings" (which assign a higher probability of $\frac{1}{2}$ or $\frac{1}{2} - \varepsilon$ to rolling a six with a loaded die) and "bad loadings" (which assign a lower probability of 0 or $\frac{\varepsilon}{2}$ to rolling a six with a loaded die). But in the previous example, the bad loadings do not assign any positive probability to rolling a six. Therefore, in the previous example, when we look at the benefit of the good loadings for rolling a six, none of that benefit is siphoned off to the bad loadings. In the present example, however, even the bad loadings assign some small positive probability to rolling a six. Thus, the bad loadings siphon off some benefit that might otherwise have been available to the good loadings for rolling a six.

Note that, in both examples, if we average over all loadings, good and bad, the probability of rolling a six is just $\frac{1}{6}$. This is always going to be the case. This, however, is not the phenomenon about which conservation of information is concerned. It is concerned about the informational costs associated with good loadings (i.e., processes that facilitate a successful outcome) and how they help increase the probability of the outcome in question. In both examples, there are five good loadings and ten bad loadings, and so the good loadings have probability $\frac{1}{3}$ and the bad loadings have probability $\frac{2}{3}$. But the good loadings contribute more probability to rolling a six in the previous example than in the present example. That is why information ends up being conserved in the one but not the other. That is the difference. For the underlying probability theory distinguishing these examples, showing in general when information is strictly conserved and when it is not, see Subsection 10.3.

These examples illustrate that information being conserved is the best that such loading schemes can do. They also illustrate that these schemes can add inefficiencies that make it more improbable to achieve a specific outcome than simply trying to achieve it directly, which in this case would be by simply rolling the original die in the hope of landing a six. That is the point of conservation of information: schemes to increase the probability of successfully achieving a certain outcome must themselves be accounted for. Left unaccounted for, such schemes become subterfuges that pretend to give something for nothing. At best, these schemes offer no probabilistic benefit. Typically, they incur a probabilistic deficit.

3. SEARCH AND THE SEARCH FOR A SEARCH (S4S)

Even though these die-tossing examples capture crucial aspects of conservation of information, it is time to step back and take a wider view of this concept. With a die, whether fair or loaded, we are apt to think of ourselves in a gambling context where the aim is to achieve a certain probabilistic outcome with the hope of benefitting financially. There are extraneous contextual details in this that are irrelevant to conservation of information. In its essence, however, conservation of information is about *search*. In the die-tossing examples, we could have described what we were doing as "searching" for a six given certain loadings of the die. (Note that even a fair die can be considered loaded in the sense that each face has the same probability of $\frac{1}{6}$).

We therefore conceive of conservation of information in terms of how information affects the success of search. Search is a very general concept that covers a wide range of activities and phenomena. It includes a gambler searching for a particular outcome at a game of chance. It includes an evolutionary process searching for a system that performs a particular function. The very notion of problem-solving falls under search: to solve a problem means searching for and then finding a solution within a space of possible solutions.

This wide-ranging use of the term *search* may in some contexts sound unfamiliar. For instance, jarring as it may seem, it nonetheless makes sense to think of a batter in baseball as searching for a way to make contact with a pitched ball to get a hit or of a golfer as searching for a way to make contact with the ball so that it lands in the hole. The underlying mathematics of search in all such cases is the same. Moreover, this conception is in keeping with a well-established literature on search.[6]. It therefore makes good sense to formulate conservation of information in terms of search.

With searches, there is always a search space, which we can denote by Ω, and there is the target that is being searched for, which we can denote by T. Obviously, if the aim is to find the target T, we want the probability of finding it to be as large as possible. In the die-tossing example, the search space comprised the six faces of the die and the target was the face showing a six. That is the target we were searching for. Success in finding the target, in this case, was a matter of rolling a die (fair or biased) that happened to land on six.

Some searches guarantee success and thus have probability 1 of finding the target T. Exhaustive search, when feasible, runs through every point in the search space to see if it belongs to the target and so will have probability 1 of finding the target. However, many search spaces are so large that an exhaustive search is impossible, and so only a tiny fraction of the full search space can actually be sampled. In general, searches can leave some uncertainty of success and thus may—and typically do—have a probability of success strictly less than 1.

Before proceeding, it should be noted that search theory—the mathematical theory of search—takes as its starting point search for fixed targets as just described. Nonetheless, search theory applies to more general targets. Search theory can accommodate false targets—targets that slow the search and distract it from the true target [6, ch. 6]. It can also accommodate moving targets—targets that have only a probability of being in a particular location at a given time and where the probability of location changes over time [9]. But measure theory (the mathematics underlying probability theory) allows for reformulating all such searches as searches for fixed targets in fixed search spaces. Such a reformulation requires working with Cartesian product spaces of the original search space and putting suitable probability measures on these product spaces. The details here need not distract us. Suffice it to say, the focus in this paper on fixed targets imposes no restriction on conservation of information as applied to targets in general.

With conservation of information, the focus is on *finding a search that with high probability finds a target* (such as finding a loading of a die that with high probability rolls a six). Let the phrase "finding a search that with high probability finds a target" sink in. It may be the most important phrase in this paper for understanding conservation of information. To talk about finding a search suggests that search is being turned in on itself—that a search is itself being searched for. Indeed, to find a search that with high probability finds a target is to engage in a higher-level search. It is to be engaged in a search for a search. This idea of a search for a search is so central to conservation of information that we abbreviate it *S4S*.

Just as a search takes place in a search space Ω, so a search for a search takes place in a higher-level search space, which we will denote with a bar above it—namely, $\overline{\Omega}$. And just as T is a target in Ω, so we will denote $\overline{T}$ as a target in the higher order search space $\overline{\Omega}$. Conservation of information unpacks the relation between the probability of T in Ω, the probability of $\overline{T}$ in $\overline{\Omega}$, and the probability of successfully finding T in Ω given that $\overline{T}$ has been found in $\overline{\Omega}$. As a preview, letting P denote the relevant probability measure, the Law of Conservation of Information then states that $P(\overline{T}) \leq \frac{P(T)}{P(T|\overline{T})}$. There are details to be worked out in this inequality since the original search space and the search-for-the-search space are separate search spaces. We therefore need to formulate the probability P so that it can simultaneously handle both search spaces. These details are addressed in Section 9.

Searching for a search makes perfect sense. In fact, we search for searches all the time, though we usually do not put it that way. If, for instance, you have a serious medical condition with several treatment options, you will want to choose the one with the highest probability of success, which is to say you want to go with the treatment that historically has led to the most recoveries. While less effective treatments might sometimes work, and even the most effective might occasionally fail, the point is to maximize the probability of success of the treatment. Effectiveness of a treatment is thus gauged in terms of the probability of recovery, and any treatment itself can be conceived as a search for recovery.

You are thus searching for a treatment that itself is a search for your recovery. The prudent choice is therefore to go with the treatment that offers the highest probability of success. In reaching your choice, you are searching for a treatment among the various treatment options with the best odds of a positive outcome. In short, your challenge in finding the right medical treatment for you is to conduct a successful search for a search. Granted, we do not usually think of or talk about such medical decisions in these terms. But the logic and mathematics underlying such decisions fit squarely within a search framework.

4. SEARCHING FOR AN EASTER EGG AND A TREASURE

The idea of searching for a search can be a mind bender, so let us consider some examples that make this idea intuitively clear. Consider an Easter egg hunt for a single egg hidden in a large field. The field is so large and the

[6] A classic in the field is [6]. Moving closer to the present, see also [7] and [8].

Easter egg is so well hidden that finding it via a blind search is extremely unlikely. In other words, simply by wandering around in the field without any information about the egg's location gives you a very small probability of successfully finding the egg. The field is so large that you simply do not have the time and resources to check all or most of the places where the egg might be hidden.

But now imagine that a friend shouts to you, "warmer, warmer, colder, cold, warmer, warmer, hot, hotter, you are burning up." As soon as your friend says that you are burning up, you look down and see the egg partially hidden in a tuft of grass. Obviously, your friend gave you information to help you in your search for the egg. Your friend thereby vastly increased your probability of finding the egg (or vastly reduced your uncertainty in finding it). But where did your friend in turn get that information that successfully directed you to the egg? Instead of giving you directions that guided you with certainty to the egg (with probability 1), perhaps the directions would only have guided you to the egg with a degree of uncertainty, such as with a probability of $\frac{1}{2}$. Or if the directions were bad (misinformation), your probability of finding the egg might even be lower than simply trying to find the egg on your own through a blind search.

The point to note is that all these ways of directing you toward the egg by shouting a sequence of instructions ("warmer," "colder," etc.) belong themselves to a search space. Some of these sequences of instructions will be productive, improving your probability of finding the egg. Others will be neutral, not helping you to do better than blind search. And others will be downright counterproductive, misdirecting you and thereby actively hindering you from finding the egg.

Consequently, your friend, in choosing one of these sequences of instructions to direct your search for the egg, had to conduct a prior search for a search. In successfully navigating the original search space to find the egg, you needed information in the form of a sequence of instructions to direct you to the egg. But in successfully navigating the higher-order search for the search (S4S) to find a search that can with high probability get you to the egg, your friend likewise needed information to search through the space of searches (the S4S space $\overline{\Omega}$). And where did that information come from? Already in this simple example we see the *information regress* that is inherent in conservation of information: as much as the original search requires prior existing information, so the search for the search requires prior existing information. The information regress problem is taken up in detail in Section 12.

You and your friend were therefore both engaged in searches: you were searching the original search space, the field with the egg; your friend was searching the S4S space, an abstract space consisting of all possible sequences of instructions for finding the egg. In this example, conservation of information says that navigating that abstract S4S space to find a sequence of instructions that helps you to successfully find the egg is no easier than navigating the original search space (the field) to find the egg.

Computer scientists, in discussing the no-free-lunch (NFL) theorems, have described how going to a higher-level search to solve a lower-level search does not gain any net benefit. Conservation of information goes beyond the NFL theorems by measuring the informational costs incurred in moving between the higher- and lower-level searches. Yet as an illustration from this earlier literature of successfully conducting a lower-level search by first successfully conducting a higher-level search, consider the passage quoted below from Joseph Culberson, published in 1998 shortly after the NFL theorems were first proven by David Wolpert and Bill Macready.[7] (Note that successfully finding an effective evolutionary algorithm—which requires finding the right representations, operators, and parameter settings to fold into the algorithm—corresponds to successfully conducting a higher-level search). Note Culberson's conclusion here, which will sound familiar given the Easter egg hunt example:

> With an evolutionary algorithm, the algorithm itself represents something of a black box with internal interactions between various operators and multitudes of parameter adjustments to the system with no clear and detailed understanding of how they interact. The researcher trying to solve a problem is then placed in the unfortunate position of having to find a representation, operators, and parameter settings to make a poorly understood system solve a poorly understood problem. In many cases he might be better served by concentrating on the problem itself. [12, p. 125]

In the Easter-egg example, this conclusion recommends simply trying to find the egg directly and not searching for a sequence of instructions that then in turn would help to find the egg. This conclusion also raises an obvious question about the Easter egg example: how did your friend come to know an effective set of instructions capable of guiding you to the egg? When parents guide children to Easter eggs by offering instructions like "warmer" and "colder," it is because they themselves hid the eggs and therefore know where they are hidden (thereby successfully resolving the higher-level search so their child can successfully solve the lower-level search). But where did your friend get his special knowledge? Without such special knowledge (information), he would

[7] For Wolpert and Macready's first work on no free lunch, see [10, 11]. For the paper by Culberson, see [12].

simply be guessing, and his instructions would be no more helpful than you simply trying to find the egg by randomly searching the field.

A variant of the Easter-egg-hunt example may provide further insight. This new example makes clearer what it means to search for a search because it provides a concrete picture of the S4S space (by contrast, a space of instruction sequences in the Easter-egg example seems more abstract). So, imagine this time that a treasure is buried on a large island. The island is so large that blindly searching the island for the treasure is infeasible—the probability of successfully locating the treasure by blind search is minuscule. But suppose next that you are given a map of treasure island for which X marks the spot where the treasure is located. That map thus gives you information for finding the treasure, raising your probability of finding it to 1.

But how did you find the map that tells you exactly where to find the treasure? Maps reside in the offices of a mapmaker—let us say the offices of Rand McNally. Rand McNally, however, is committed to having a comprehensive collection of maps. Thus, for every location on treasure island, it would have a map with an X marking that location. In that case, finding the right map by chance is just as hard (probabilistically speaking) as finding the treasure by chance. But the situation could easily be worse. Rand McNally could also have maps of the island with multiple X's where the treasure might be located, the vast majority of these being false targets. It could even have maps that describe entire paths on which the treasure might be located. Moreover, as mapmaker to the world, Rand McNally would have maps of many other land surfaces, all of which would need to be sorted through before even getting to maps of treasure island. Most of the maps at the Rand McNally offices will not help you find the treasure. Most will distract or even misdirect you. Only a tiny proportion of its maps will help you successfully find the treasure. Conservation of information says that finding such a map (and thereby completing a successful search for a search) will be so difficult that at the end of the day you are better off simply doing a blind search of the island for the treasure.

Rand McNally's offices, with its full array of maps, provide a concrete picture of the S4S space. This example also helps clarify why strict conservation of information is the best that we can do in the search for a search, and that in fact the search for a search may incur inefficiencies that undercut the probability of finding the target (in this case, the treasure), so that information funneled through a search for a search ends up being unconserved. In general, the S4S space (in this case, the offices of Rand McNally) will be much larger—in the sense of containing many more possibilities to be sorted through—than the original search space. The original search space here consists of every location on the island where the treasure might be. But the Rand McNally offices include an immense number of maps that, with respect to all the possible places where the treasure might be located, is even more fine-grained than the possible locations on the island for the treasure. Searching through these maps (each map constituting a search) thus ends up being more information intensive than simply searching the island directly.

This increase in complexity of the search for a search over the original search accounts for why strict conservation of information is the best we can hope for in searching for a search that successfully finds the original target. The search for a search is a search for a higher-level target consisting of searches that with high probability find the original target. Locating such a higher-level target and then using a search from it to try to find the original target can strictly diminish the probability of finding it below simply searching for the target directly, thus doing worse than strict conservation of information.

5. THE BASIC FRAMEWORK FOR CONSERVATION OF INFORMATION

In light of these examples, let us now turn to the basic framework for conservation of information. With conservation of information, we always start with a *baseline search* that has a *baseline probability* of success in finding a target. This baseline probability is typically very small, and the baseline search is typically a uniform random search or some other form of blind search that has no special information about the location of the target, and thus provides no advantage for finding it. In general, it can be any search that is at least as good as a blind search since we need never do worse than blind search. The challenge is to do better than a baseline search when its probability of locating the target is minuscule.

In most practical cases involving conservation of information, the baseline probability is so small that the baseline search is extremely unlikely to succeed. This is different from larger baseline probabilities, in which running baseline searches are likely to succeed, especially if they are run repeatedly. In that case, there is little incentive to search for better searches. Conservation of information becomes interesting when the baseline search is a matter of finding a needle in a haystack. Conservation of information then tracks the information needed to find the needle. If a search requires finding a needle in a haystack without any outside aid, then the search is destined to fail. The challenge, then, is to discover and exploit information that makes finding the needle (the target) easier (more probable).

To that end, we need an *improved search* with an *improved probability* of successfully finding the target. Running that improved search will then help us to find the target with much higher probability than the baseline probability. In most instances, the improved probability will be close to or even equal to 1, whereas the baseline probability is close to 0. The question then becomes how we were able to secure the improved search with its improved probability. To that end, we needed to perform a *search for a search*, or S4S. Associated with the S4S is an *S4S probability* of successfully finding a search that can then find the original target with a probability at least as large as that of the improved probability.

In conservation of information, we therefore always have to keep track of three searches and their three corresponding probabilities of successfully locating targets: (1) the baseline search with a very small baseline probability of successfully finding the original target; (2) the improved search with a much larger improved probability of successfully finding the original target; and (3) the search for the search (S4S) on a higher order search space, the S4S space, with an S4S probability of successfully finding a search that in turn finds the original target with a probability at least as large as the improved probability. This S4S space therefore includes a higher-order target consisting of all searches that succeed in finding the original target with a probability at least as large as the improved probability. With a tiny baseline probability and a large improved probability, conservation of information then establishes that this S4S probability will itself be tiny.

Some standard notation for the baseline, improved, and S4S probabilities helps to make clear how these probabilities work together in conservation of information. We use p to denote the baseline probability. Usually p is very very small: close to 0. Next, we are given an improved search with probability q of search success, where q is much bigger than p, which may be denoted by $q \gg p$. Often q is close to 1 or even identical to 1. And finally, we are given an S4S probability r of finding a search whose performance in finding the original target is at least as effective as that of the improved search. Effectiveness here is gauged in terms of probability, and so any search discovered by such a search for a search will have probability at least q of finding the target.

As in Section 3, if we let T denote the original target in the original search space, then the search for a search tries to locate a higher-level target $\overline{T}$ (consisting of all searches that meet or exceed the performance of the improved search) in the S4S space. Any search taken from this higher-level target will then have probability at least as large as q of finding the original target T. In this case, r will denote the probability of the higher-level target $\overline{T}$. This higher-level target denotes all the possible improved searches that increase search performance from the baseline probability p to at least that of the improved probability q. In the die-tossing example that opened this paper in which the aim is to roll a six, the higher-level target is all weighted dice that assigned a probability of $\frac{1}{2}$ to two faces, one of which was a six. There were five such weighted dice (the higher-level target) among a total of fifteen (the entire search-for-a-search space).

Conservation-of-information theorems then establish that the search for an improved search that succeeds with probability q or greater has a probability of success no greater than $\frac{p}{q}$. Given our basic framework, the search for this improved search is denoted as having probability r of success, or equivalently, the probability of the higher-level target $\overline{T}$ in the S4S space is r. Conservation-of-information theorems then show that $\frac{p}{q} \geq r$. In case r exactly equals $\frac{p}{q}$, we say information is conserved. In case r is strictly less than $\frac{p}{q}$, we say information is unconserved.

Note that because q is a probability, it cannot exceed 1, and so $\frac{p}{q}$ will always be greater than or equal to p. If q actually equals 1, the probability r of successfully conducting a search for an improved search will be less than or equal to p. On the other hand, if q is strictly less than 1, $\frac{p}{q}$ will be strictly greater than p (denominators less than 1 increase fractions of positive numbers), allowing for r also to be greater than p. Conservation of information then says that finding an improved search with probability q of successfully finding the original target itself has probability r that is less than or equal to $\frac{p}{q}$. This means that the one-two punch between the search for a search and the improved search working in tandem cannot make finding the original target any more probable than the baseline search with its original baseline probability p. Conservation of information is the best that these probabilities can do. In general, a search for a search may have a probability r that is less than $\frac{p}{q}$. The probability bound $\frac{p}{q}$ ensures that finding the original target via a search for a search cannot improve on the original baseline probability p.

Let us now connect these numbers to the die-tossing example of Section 2 in which any two faces can be loaded so that each lands with probability $\frac{1}{2}$. Note that for purposes of exposition, these probabilities are much larger than typically arise with conservation of information in real-world applications. But the math here cares only about whether the baseline probability p is strictly less than the improved probability q. With this die-tossing example, the baseline probability of landing six is $p = \frac{1}{6}$, and the improved probability of landing six is $q = \frac{1}{2}$. Moreover, the S4S probability is, as we showed, $r = \frac{1}{3}$, which equals $\frac{p}{q}$. Consequently, in this example information is conserved.

Although the probabilities just described capture the mathematics in conservation of information, the existing conservation of information literature typically transforms these probabilities into information measures by applying the negative logarithm to the base 2. Thus, for T, the target in the original search space with baseline probability p, and for $\overline{T}$, the higher-level target in the search-for-a-search space where each search in it confers an improved probability of at least q, the S4S probability of finding $\overline{T}$ is r, and so by conservation of information $r \le \frac{p}{q}$. These probabilities can then be denoted respectively by $P(T) = p$, $P(T|\overline{T}) = q$, and $P(\overline{T}) = r$ (the precise meaning of this notation will become clear in Section 9). Conservation of information then says that $r \le \frac{p}{q}$ or equivalently that $P(\overline{T}) \le \frac{P(T)}{P(T|\overline{T})}$.

If we now transform these probabilities by the negative logarithm to the base 2, we turn probabilities into bits, which typically are fractional rather than integer-valued (for instance, the probability $\frac{1}{10}$ corresponds to $-\log_2(\frac{1}{10}) \approx 3.322$ bits). Letting P denote a probability measure and I the corresponding information measure, we thus have:

$$I(T) = -\log_2 P(T) = -\log_2(p),$$
$$I(T|\overline{T}) = -\log_2 P(T|\overline{T}) = -\log_2(q), \text{and}$$
$$I(\overline{T}) = -\log_2 P(\overline{T}) = -\log_2(r).$$

Thus, the key probabilistic inequality of conservation of information (namely, $r \le \frac{p}{q}$) becomes, in information-theoretic terms, $-\log_2(r) \ge -\log_2(p) + \log_2(q)$, or equivalently $I(\overline{T}) \ge I(T) - I(T|\overline{T})$. What makes this inequality interesting is that when the search for a search is successful, the probability $P(T|\overline{T})$ will be large (close to 1), making the corresponding information measure $I(T|\overline{T})$ small (close to 0). Consequently, the information in a successful search for a search—namely, $I(\overline{T})$—will be roughly the same as the information in a successful search—namely, $I(T)$—implying that the search for a search did not help resolve the original search.

In the conservation of information literature, each of these information measures has a name: (1) for the baseline probability $P(T) = p$, the corresponding information measure $I(T)$ is called the *endogenous information*; for the improved probability $P(T|\overline{T}) = q$, the corresponding information measure $I(T|\overline{T})$ is called the *exogenous information*; and for the S4S probability $P(\overline{T}) = r$, the corresponding information measure $I(\overline{T})$ is called the *active information*, or equivalently the *added information.* Moreover, in this literature, the baseline search is called the *null search* or *blind search* and the improved search is called the *alternative search*, but the *search for a search* keeps its name unchanged.

Endogenous information captures the inherent uncertainty in finding the target T. Exogenous information captures the uncertainty that is left once an improved search has been implemented. Active/added information captures the uncertainty in procuring an improved search. In information-theoretic terms, conservation of information can therefore be stated as follows: **active information is always at least as great as the difference between endogenous and exogenous information**.[8]

In symbols, as we've already noted, this is just $I(\overline{T}) \ge I(T) - I(T|\overline{T})$. So stated, this formula may seem a bit abstract. In practice, however, it admits a helpful simplification. Typically, $P(T|\overline{T}) = q$ is close to 1. Let us say it is at least $\frac{1}{2}$. In that case, the negative logarithm to the base 2 of this probability will be less than or equal to 1 bit. In other words, $I(T|\overline{T}) \le 1$. Because $P(T) = p$ will typically be very small and close to 0, it follows that $I(T)$, which we can call b, will be a large number of bits (if, for instance, $p = 10^{-150}$, then $b \approx 500$). Accordingly, $I(\overline{T})$, the active information, would be greater than or equal to $b - 1$. Generally speaking, in situations where conservation of information most readily applies, if the endogenous information is b bits, the active information will therefore be roughly b bits or more (often considerably more).

In applications of conservation of information, active information (corresponding to the S4S probability) is always the key point of interest—how did it come about, how does it successfully facilitate a given search, and what is its ultimate source? Inspiration for the term *active information* as used in relation to conservation of information can be credited to the physicist John Polkinghorne, who introduced it as a form a top-down causality capable of doing teleological work, though he used the concept more to advance non-reductionism than teleology per se.[9]

Information engineers may find this information-theoretic formulation of conservation of information to be intuitively more appealing than the corresponding probabilistic formulation. Despite this, the two are logically equivalent. That said, the probabilistic formulation seems more basic in that information metrics are defined in terms of probability metrics, and not vice versa. In any case, having thus far focused mainly on the probabilistic formulation, I will continue to focus on it in the sequel, if only for ease of exposition. The exception will be Appendix 2, where I review several earlier conservation-of-information theorems, which were formulated explicitly in information-theoretic terms. As it is, I leave it to the reader to translate the probabilistic formulation into the equivalent information-theoretic formulation (and vice versa, as the case may be).

[8] This information-theoretic formulation of conservation of information first appeared in the scientific literature in [13].

[9] "The portfolio of causes that bring about the future is not limited solely to the description offered by a methodologically reductionist physics and framed only in terms of the exchange of energy between constituents. Instead, the concept of causal influence can be broadened at least to include holistic effects of an informational, pattern-forming kind. One might call this top-down form of causality 'active information'." Quoted from [14, p. 35].

6. TARGETS IN SEARCH

How widely does the idea of targeted search apply? Does it apply only to situations in which an intelligent agent is actively seeking to achieve some purpose? Or can it apply more generally to natural processes that without any evident purpose arrive at some salient outcome? The use of the word "target" in relation to search therefore requires some clarification. The language of searching for targets can seem inherently teleological, presupposing intentional guidance by a purposive agent. If targeted search were inherently teleological, the use of conservation of information in intelligent design would be severely undercut. Any discovery of intelligence in the search process would then be the conclusion of a question-begging argument where intelligence is concluded only because it was first presupposed.

In fact, searching for targets need not presuppose teleology and can be metaphysically neutral about the types of causes at play in search. Obviously, in human contexts, where search is for humanly invented products brought about by humanly invented processes (e.g., building a better mousetrap by building a better mouse-trap making machine), the targets in question are chock-full of teleology or purpose. Indeed, when we talk about targets in these contexts, seeking and attaining targets results from a conscious human intelligence aiming at the targets, trying to bring them about for a purpose.

What happens, however, when we try to understand targets in the context of nature? Scientific naturalists see nature as self-contained, impervious to outside (supernatural) influence and operating by unbreakable natural laws. For them, purpose is not inherent in nature but a byproduct of nature only after nature evolves living forms capable of exhibiting purpose. Use of the word "target" in the natural sciences can therefore suggest a preordained outcome, such as a specific biological or cosmological structure, as though nature were (quasi-intentionally) directed toward achieving it. To scientific naturalists, this perspective is incompatible with evolutionary theory, which posits undirected processes as sufficient to build complexity via incremental changes (such as natural selection acting on random variations). To frame conservation of information around achieving specific targets is thus seen as illicitly projecting purpose onto an otherwise purposeless nature.

Notwithstanding such criticisms, targets do not need to presuppose purpose. One way to see this is to note that the very term "evolutionary search" appears widely in the biological evolution and evolutionary computing literature (just key in this term at Google Scholar). Even as hard-core a Darwinist and scientific naturalist as the late Daniel Dennett put search at the heart of evolution, referring to "the ultimately *algorithmic search processes* of natural selection" [15, p. 144, emphasis added]. But what exactly is such an evolutionary search searching for? In biology, the search space is a space of possible biological configurations, and what is being searched for in that space is some biologically distinguished subset, such as polypeptides that fold into functional proteins. Such subsets form the targets of evolutionary search. Yet they do not presuppose purpose or teleology.

In the same vein, consider the following quote by self-organizational theorist and scientific naturalist Stuart Kauffman, for whom Darwinian processes are necessary, though by no means sufficient, for evolution. In his view, Darwinian theory leaves a crucial question unanswered—namely, the structure of the fitness landscapes on which evolution depends. For Kauffman, evolution is a form of search that requires properly structured fitness landscapes before it can go anywhere. As he puts it:

> For the 3.45 billion years of life, simple and complex organisms have been adapting, accumulating useful variations, climbing up fitness landscapes toward peaks of high fitness. Yet we hardly know what such fitness landscapes look like or how successful an *evolutionary search process* is as a function of landscape structure. Landscapes can be smooth and single peaked, rugged and multipeaked, or entirely random. Evolution searches such landscapes using mutation, recombination, and selection. [16, p. 161, emphasis added]

Note the explicit reference to an "evolutionary search process." The targets here that are being searched for are those places where the fitness landscape is high. Also implicit in this quote are higher-level searches (the search for the search, or S4S): evolutionary theory is for him not just a theory about how to search a given biological configuration space by means of a given fitness landscape but more so about how to search among different types of fitness landscapes (smooth, rugged, random, etc.) for those that allow a successful form of evolution. In this connection, Kauffman writes:

> Life uses mutation, recombination, and selection. These search procedures seem to be working quite well. Your typical bat or butterfly has managed to get itself evolved and seems a rather impressive entity... Mutation, recombination, and selection only work well on certain kinds of fitness landscapes, yet most organisms are sexual, and hence use recombination, and all organisms use mutation as a search mechanism... Where did these well-wrought fitness landscapes come from, such that evolution manages to produce the fancy stuff around us? [17, p. 19]

Kauffman admits this to be a huge open question: "No one knows." [17, p. 18]

Origin-of-life researcher Robert Hazen, like Dennett and Kauffman, displays a naturalistic, non-purposive view of targets in his proposed Law of Increasing Functional Information. At the heart of this law, as he frames it, is calculating the proportion of functional things within a wider reference class of things that are functional as well as non-functional. The functional things obviously form a target within the wider reference class, which with some further assumptions allows a probability to be assigned to the target. Hazen appears to be in the early stages of developing the Law of Increasing Functional Information, which he understands in broadly evolutionary terms.[10] Yet the point to recognize is that the notion of a target is central to Hazen's law. Indeed, the notion is essential to all evolutionary theories regardless of whether they explicitly use the term or only tacitly engage the underlying concept.

Many targets cannot be connected to human or other clearly identified purposive agents. In general, such targets—call them "natural" or "neutral"—exhibit some scientifically significant feature that calls for further inquiry and explanation. In biology these targets typically satisfy some objectively given functional specification (such as the ability to replicate). Natural targets like this fall under what philosophers call "natural kinds."[11] Conservation of information legitimately applies to all such targets. To the degree that these targets are improbable, their attainment needs to be explained. Yet the targets themselves need not presuppose teleology or purpose. Without prejudging the role of chance or intelligence in the attainment of targets, conservation of information simply sorts through the informational requirements for their attainment. **Conservation of information is an accounting principle**. It is concerned with ensuring that the numbers on a balance sheet add up. It is not concerned with where the numbers ultimately come from.

[10]See the following video where Hazen sketches the Law of Increasing Functional Information: https://www.youtube.com/watch?v=lepxTr9zKDc (last accessed December 12, 2024).

[11]Natural kinds are categories or groupings in nature that reflect real, objective divisions rather than arbitrary or human-constructed classifications. Natural kinds are thought to have intrinsic properties that define their identity, making them indispensable to scientific explanation and inquiry. For instance, in chemistry, gold and water are considered natural kinds due to their objectively given characteristics (atomic structure and molecular composition). Philosophically, natural kinds underpin discussions about realism, essentialism, and the nature of scientific classification. They are central to debates about whether such kinds exist independently of human cognition or are merely convenient tools for organizing our observations of the world. This concept plays a critical role in fields like metaphysics, philosophy of science, and epistemology, addressing how we structure and understand the natural world. Natural kinds help us to make sense of how targets in search need not presuppose human or other intelligences. For more on natural kinds, see the *Stanford Encyclopedia of Philosophy*, s.v. "natural kinds," https://plato.stanford.edu/entries/natural-kinds (last accessed January 9, 2025).

Despite the case just made for the validity of targeted search in modeling evolutionary biology, I need to say something more to critics who object to evolution being conceived as search and thus reject that targets are relevant to understanding evolution. I have just shown that this rejection is untenable as targeted search falls squarely among the challenges that evolutionary biology must face. Nonetheless, it is worth considering how some scientists try to circumvent this challenge. I will focus on H. Allen Orr, an evolutionary biologist who specializes in population genetics. In reviewing my 2002 book [18], Orr wrote:

> Darwinism is not trying to reach a prespecified target. Darwinism, I regret to report, is sheer cold demographics. Darwinism says that my sequence has more kids than your sequence and so my sequence gets common and yours gets rare. If there is another sequence out there that has more kids than mine, it'll displace me. But there is no pre-set target in this game. (Why would evolution care about a pre-set place? Are we to believe that evolution is just inordinately fond of ATGGCAGGCAGT. . . ?) Dembski can pick a prespecified target, average over all fitness functions, and show that no algorithm beats blind search until he's blue in the face. The calculation is irrelevant. *Evolution is not searching for anything and Darwinism is not therefore a search algorithm.* The bottom line is not that the NFL theorems are wrong. They are not. The bottom line is that they ask the wrong question for what Dembski wants to do. More precisely, the proper conclusion is not that the NFL theorems derail Darwinism. The proper conclusion is that evolutionary algorithms are flawed analogies for Darwinism.[12]

In *No Free Lunch*, the book under review, I examined the no-free-lunch theorems of Wolpert and Macready. Clearly, the title of that book alludes to these theorems. Yet nowhere in the book did I claim that these theorems refute Darwinian evolution. The theorems, as I used them, set the stage of a critical reexamination of Darwinism's core tenets. No free lunch states that averaged over fitness landscapes, no search outperforms any other, and so all searches, when averaged, are equivalent to blind search. But the fact is that in specific situations, some searches do outperform others. The point of my

[12]H. Allen Orr, Review of No Free Lunch: Why Specified Complexity Cannot Be Purchased Without Intelligence, *Boston Review*, June 2002: https://www.bostonreview.net/articles/h-allen-orr-review-no-free-lunch (last accessed February 15, 2025). Emphasis added.

book *No Free Lunch* was to analyze the probabilistic hurdles to finding the better-performing searches and then argue that these hurdles posed insuperable obstacles to naturalistic theories of evolution, thereby creating room for intelligent design. Twenty years ago, that was the problem I was trying to resolve. Yet I did not back then have a complete solution to it as I do now through subsequent work on conservation of information culminating in this paper. The broad outlines of that solution, however, were evident even back then.

Orr in his review presents a straw-man argument. For him, targets prespecify organisms with such overprecision that they render search for these targets biologically ridiculous. It is as though, in his words, evolution could be "inordinately fond of ATGGCAGGCAGT," but this is a straw man. No one thinks that evolution can insist on bringing about some particular sequence of nucleotides like this, especially since genomes always tolerate many alternates and substitutions capable of maintaining structure, function, and viability. In biology, the targets that are the focus of conservation of information are biologically significant functions whose origin requires explanation. At the top of the list are the core life functions such as DNA replication, metabolism, homeostasis, protein synthesis, gene regulation, etc.

For Orr to characterize Darwinism as "sheer cold demographics" is therefore simplistic and misleading, even on Darwinian grounds. Demographics is about the reproductive output of organisms considered collectively. Crucially lacking from demographics is what features of organisms allow them to exist and reproduce at all and how these features arose in the first place. Additionally, demographics is indifferent to the range of features capable of evolving. Demographics allows a world of extremely simple replicators that never show any interesting evolution and that are unimpressive given the actual complexity and diversity of life that we witness. Demographics by itself cannot distinguish between Stone-Age technology and modern digital technology. Demographics cannot explain why life is as high-tech as we find it.

I have made here what I take to be a solid case for conceiving of evolution as a targeted search, defining targets in a way that is compatible with naturalism or materialism. Nonetheless, when I have made this case in print and before others, I still get pushback from evolutionists who refuse to see evolution in these terms.[13] For them, treating evolution as a targeted search is unhelpful and misguided. Accordingly, they will argue that while metaphors like "search," "target," and "fitness landscape" may be heuristically useful, they carry unintended teleological baggage that distorts how evolutionary biologists actually understand and model change. For them, mutation, far from being an exploratory mechanism, is a stochastic byproduct of imperfect replication, often maladaptive in the short term. Moreover, fitness landscapes remain largely intractable in practice—high-dimensional, dynamic, and unmeasurable—so their use risks oversimplifying the complexity of real evolutionary processes. Perhaps most significantly, positing targets is said to imply evolutionary directionality, but evolution is, according to them, more accurately understood as contingent traversal through biological configuration space, with no fixed destination and only local criteria for success reflected in population-level outcomes.

Despite such concerns, treating biological evolution as a form of targeted search is both conceptually sound and scientifically fruitful—particularly when the concepts of "search" and "target" are used in a metaphysically neutral way. A target need not be consciously chosen or represent a global optimum; it may simply refer to a grouping of functionally significant outcomes, such as the emergence of reproductive or metabolic strategies, that are statistically rare and thus probabilistically nontrivial to attain. The use of fitness landscapes, though idealized, captures essential features of evolutionary dynamics: local improvements, trade-offs, and ruggedness in adaptive potential. The combined action of mutation and selection is readily conceived as an algorithmic search process that differentially samples the underlying biological configuration space. Indeed, it is no accident that evolutionary computing, which is all about search, takes its inspiration from evolutionary biology. Moreover, the need to explain how core biological functions arose—functions without which reproduction and thus demographics are impossible—demands a framework like targeted search that can account for probabilistic barriers and informational constraints.

Turning the tables, we may now ask: If biological evolution is not search, what is it? It is not enough to say that it is a non-teleological process subject to sheer cold demographics. If not search, what is the model of evolution that explains all the cool stuff we see in biology? H. Allen Orr, with his focus on demographics, might say that the model is population genetics. Population genetics, however, merely offers descriptive statistics for quantifying how gene frequencies fluctuate in populations subject to various environmental influences. But where do we get the genes that do all the interesting stuff that makes the demographics of living things worth studying in the first place? Targeted search provides a well-defined model for answering this question. Those who deny that evolution is a targeted search offer no model in its place. Yet without such a model, there is, strictly speaking, no evolutionary theory, which in turn would mean that evolution cannot rightly be regarded as a science.

[13]See, for instance, my 2014 talk on conservation of information at the University of Chicago: https://www.youtube.com/watch?v=MN74Vn-R5fg (last accessed May 21, 2025).

To sum up, far from prejudicially imposing teleology on biology, conservation of information quantifies what any process, including a non-teleological one, must overcome to achieve novel biological information. The features that organisms need to exist and flourish can rightly be regarded as targets. Moreover, any theory of evolution worth its salt must be able to explain how these targets can be attained. Many evolutionists, such as Kauffman and Hazen, are willing to treat evolution as a search and credit natural selection as vital to it. In rejecting that evolution is a search, Orr mistakenly requires any targets in biology to be prespecified or preset, as though they were artifices humanly imposed on biology. But these targets exist as a matter of biological reality. Life must achieve these targets to exist and thrive. Indeed, targeted search makes good sense for understanding biological innovation, especially the origin of core life functions.

7. SPECIFIED COMPLEXITY

Readers familiar with specified complexity may wonder what its connection is to conservation of information and whether the two might simply be different ways of referring to the same underlying concept. As it is, the two are related but also distinct. In this section, I want to give a thumbnail sketch of specified complexity and indicate how it connects to targeted search in conservation of information. This will require a brief review of specified complexity. Specified complexity and conservation of information together form the twin mathematical pillars on which the theory of intelligent design is founded, so it is important to understand the connection.

The problem that specified complexity is meant to resolve is this: given some event or object (and by extension some target answering to either) for which we do not know exactly how it came about, what features of the event or object could lead us rightly to think that it was the product of a designing intelligence? This question asks us to engage in effect-to-cause reasoning. In other words, we see an effect and then we must try to determine what type of cause produced it.

The problem that specified complexity attempts to resolve therefore differs from detecting design through cause-to-effect reasoning. In cause-to-effect reasoning, we witness the activity of a known cause and then track its effect. Thus we may see someone take hammer and chisel to a piece of rock and then watch as an arrowhead is produced. Detecting design in such a case is straightforward because we know that the person shaping the rock is an intelligent agent, and we see this agent in real time bring about an artifact, in this case by inputting information into the rock to form the arrowhead.

With specified complexity, because the reasoning is from effect to cause, we are not handed a smoking gun in the form of an intelligent agent who is witnessed to produce a designed object. Rather, we are simply given something whose design stands in question (such as a chunk of rock) and then asked whether this rock has features that could reasonably lead us to think that it was the product of design. Thus, if we were simply presented with the rock but did not see it being crafted by an arrowhead maker, our task would be to infer design simply from the rock exhibiting the shape of an arrowhead.

The key question specified complexity raises can therefore be framed as follows: Given an event whose precise causal story may be unclear (no smoking guns, no video cameras running, no direct evidence of design), what about it could convincingly lead us to conclude that it is the product of a designing intelligence? To keep things simple, we will focus on events, thereby tacitly identifying physical or digital items with the events that produce them and likewise targets with the events by which they are attained. To further simplify things, let us look at one particular type of example that captures what is at stake in this question—namely, the Search for Extraterrestrial Intelligence, or SETI.

SETI researchers find themselves at the receiving end of the classic Shannon communication diagram shown in Figure 1. This diagram, which figures centrally into Claude Shannon's 1949 book tracks information from a source to a receiver in the presence of noise [4, p. 7]. In formulating this diagram, Shannon assumed that the information source was an intelligent agent. But that is not—strictly speaking—necessary. It could be that the source of the signal is unintelligent, such as a stochastic or deterministic process. Thus, a quantum device generating random digits might instead be at the source.

Receiving signals at the far end of this diagram, SETI researchers want to know the type of cause responsible for the signals at the other end. To keep things simple, yet without loss of generality, let us assume that all the signals that the SETI researchers receive are bitstrings, that is, sequences of 0s and 1s (given enough bits, they can represent any signals—as when a fixed number of bits represent keyboard symbols). There is a lot of radio noise coming in from outer space. There are also a lot of humanly generated radio signals that SETI researchers need to exclude. So, given a radio signal in the form of a bitstring that is verifiably from outer space—and thus not humanly generated—how can we tell whether it is the product of intelligence?

The problem here is not as in the film *ET*, where an embodied alien intelligence actually lands on Earth and makes itself immediately evident. Rather, as in the film *Contact* (based on a novel of the same name by Carl Sagan), all the SETI researchers have to go on is a signal, and the question is whether its source is intelligent or non-intelligent. Any such intelligence is not immediately evident. Rather, such an intelligence is, as we might say, mediately evident. In other words, the

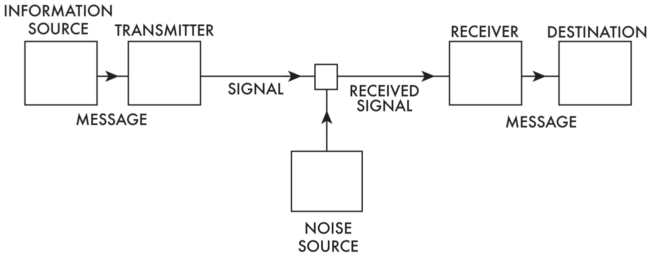

Figure 1: The classic Shannon information diagram. doi:10.5084.BIO-C.2025.2.f1

intelligence is mediated through the signal, the medium of communication.

In SETI research, the challenge is therefore to determine whether the source of a bitstring is intelligent *at all*. The presumption is that the source is unintelligent until proven otherwise. That is the default explanation. What, then, about a bitstring received from outer space could convince us otherwise? As it is, no such bitstring bearing unmistakable marks of intelligence has yet to be observed. SETI is therefore a research program that to date has zero confirmatory evidence. But that does not invalidate the program. The deeper question that SETI raises—and that legitimizes it as a research program—is the counterfactual possibility that it might pan out: What about such a bitstring would convincingly implicate an intelligence if it were observed?

One obvious immediate requirement for any such bitstring to implicate intelligence is improbability. In other words, the event in question must be highly improbable or, equivalently, it must have small probability. By contrast, events that are highly or even moderately probable are readily explained by chance. For instance, we do not question whether a coin that has landed heads resulted from anything other than chance. With the coin having so large a probability as $\frac{1}{2}$ of landing heads, we think that chance is a perfectly acceptable explanation for how it landed and we do not look for alternative explanations that invoke design.

What does it mean for an event to have small probability? The answer depends on the number of opportunities for the event to occur, or what are known as its *probabilistic resources* [19, ch. 4]. For instance, getting 10 heads in a row has a probability of roughly 1 in 1,000. That may seem small until one considers all the people on earth tossing coins. Factoring in all those coin-tossing opportunities shows 10 heads in a row to be not at all improbable from the vantage of human experience. Indeed, if you get out a coin right now and start tossing it, you will be likely to see 10 heads in row in an hour or two (or more quickly the faster you can toss the coin).

But what about 100 heads in a row? Getting that many heads in a row has a probability of roughly 1 in 10 raised to the 30^{th} power, or 1 in a million trillion trillion. If all the humans that have ever lived on earth did nothing with their lives but toss coins, they should never expect to see that many heads in a row. It is estimated that at most 100 billion people have ever lived. Each of them would need to toss a coin 10 billion billion times for all of them jointly to have about an even chance of seeing 100 heads in a row. The improbability of 100 heads in a row therefore raises doubts about its chance occurrence—doubts that do not arise with a single head or even 10 heads in a row.

So, one requirement for a bitstring from outer space to qualify as detectably designed is for it to be improbable in relation to any plausible chance processes that might produce it. In particular, any bitstring received must be sufficiently long (a handful of bits will not do). But there is also another requirement for such a bitstring to be detected as designed. Brute improbability is not enough. Highly improbable things happen by chance all the time. If a bucket of marbles is spilled on your living room floor, their precise arrangement will be highly improbable. But if those marbles arrange themselves to spell “Welcome to our home,” anyone will instantly know, because of the pattern in this arrangement, that those marbles did not randomly organize themselves in this way. Rather, it will be clear that an intelligence is behind that pattern of marbles, and this will be clear even if the precise causal story by which the intelligence acted remains a matter of ignorance.

What is it about a pattern like this that, in conjunction with improbability, leads us to infer design? The short answer is that the pattern needs to be *recognizable*. In SETI research, for instance, where we try to detect design at the receiver of the Shannon communication diagram, we need a bitstring that is at once improbable and also exhibits a recognizable pattern (similar to marbles spelling out the words “Welcome to our home”). Any long bitstring will be improbable. But the overwhelming majority of them will not be detectably designed. Our challenge as SETI researchers, therefore, is to elaborate what qualifies as a recognizable pattern that in the presence of improbability marks a bitstring as detectably designed. In the film *Contact*, the recognizable pattern that convinced the radio astronomers that contact had indeed been established was a long sequence of prime numbers.

In practice, we detect design without attempting to measure or quantify the recognizability of patterns by which it is detected. Usually we just experience an *aha moment* when faced with a pattern that we intuitively grasp to be improbable and recognizable. It is as when Sherlock Holmes infers design by seeing isolated pieces of evidence all suddenly converge to solve a mystery. It is as when a person looks at what initially seems like a random inkblot but suddenly notices a familiar object that makes the design unmistakable and thus removes any doubt about these being merely random splotches of ink.

A rigorous theory of design detection requires that we define, in precise mathematical terms, what it is for a pattern to be recognizable. Based on algorithmic information theory (AIT, also known as Komogorov complexity theory) [5, p. 466], the recognizable patterns are those with short description lengths. These patterns are called *specifications*. Moreover, based on Shannon's information theory, a small probability p can be recast as an information measure $-\log_2(p)$, which signifies greater complexity (i.e., more bits) the smaller the probability p. The term *specified complexity* therefore denotes a recognizable pattern ("specified") conjoined with a small probability ("complexity"). In fact, specified complexity can be formulated as a full-fledged information measure defined by subtracting Kolmogorov information from Shannon information [19, ch. 6].

Though there are many details here about specified complexity that have been left unsaid, we are nonetheless now in a position to connect this concept to targeted search and conservation of information. Suppose, then, we are given an event that exhibits specified complexity. In the logic of the design inference, specified complexity is a reliable marker of intelligence. Accordingly, this event will trigger a design inference, entitling us to conclude that the event is the product of intelligent design. At the same time, the event can also be conceived as the attainment of a target by matching a recognizable pattern or specification. In either case, we are dealing with something intelligently designed—an event that is the product of design, or equivalently a target that is attained by means of design.

But as developed in this paper, conservation of information is about what happens with a search process when its probability of success starts out small and then gets increased. Left to itself, a target T may have small probability p. But what if the process that brings about T receives an informational boost that raises its probability to q where q is so large that it cannot rightly be regarded as a small probability? Suddenly T is now no longer improbable and therefore now no longer exhibits specified complexity. The small probability by which T formerly exhibited specified complexity has now evaporated.

This is the standard move by Darwinists to defeat specified complexity in evolutionary biology. Looking to natural selection as a designer substitute, Darwinists see in the Darwinian mechanism of natural selection acting on random variations a way to amplify the probabilities by which complex biological systems arise (this point becomes especially clear in the next section as we consider Richard Dawkins' METHINKS IT IS LIKE A WEASEL example). This blanket approach to invalidating specified complexity across all of biology fails: there are many good examples of specified complexity in biology that confirm design for actual biological systems even when factoring in Darwinian and other naturalistic evolutionary mechanisms. Consider, for example, some of the irreducibly complex biochemical systems analyzed by Michael Behe.[14]

But what if specified complexity could, whether in specific instances or across the board, be invalidated because probabilities previously thought to be small were in fact much larger? Conservation of information would then counter such attempts to defeat specified complexity. It would do so not by disputing earlier probability calculations, though they might be disputable. Rather, it would do so by showing that even if the probability could be increased sufficiently so that it is no longer small—as by enhancing a search process through natural selection—that very enhancement incurs a probability cost that remains unexplained. Enhancing the search therefore raises its own difficulties that do nothing to relieve the burden of having to explain an otherwise small-probability event.

Note that targeted search as it arises in conservation of information does not presuppose the full apparatus of specified complexity (an apparatus that is technically demanding). Conservation of information is simply concerned with the informational costs associated with increasing the probability of a target's successful search, however that target is defined or identified. Specified complexity can be grafted onto this conservation of information framework so that the targets considered are those that exhibit specified complexity. But conservation of information can make sense of targets that are not specified in the sense required by the design inference (though often they are so specified). That said, conservation of information will always apply to targets that are complex or improbable (regardless of whether they are also specified in some precise technical sense). That is because non-complex or probable targets require no special explanation of why a search for them can succeed—if they are probable, they are likely to be found.

Taken together, specified complexity and conservation of information form a potent two-part challenge to facile dismissals of design: Either (1) admit that an event exhibits specified complexity and is therefore designed, or (2) deny that the event exhibits specified complexity because its small probability undergoes a drastic increase, only to be obligated by the Law of Conservation of Information to explain that drastic increase.

8. THE DISPLACEMENT FALLACY

The discovery of conservation of information did not start with proving mathematical theorems. Rather, its

[14] See [20] for examples of irreducible complexity. For connecting irreducible complexity with specified complexity, see [19, ch. 7].

discovery came from repeatedly noticing how efforts to account for the success of searches whose odds of success were otherwise hopeless always smuggled in information that wasn't properly accounted for. One hole was filled, but only by digging another, and so a new hole now in turn needed to be explained. This failure of explanation became especially evident in the evolutionary literature. Darwinian approaches to biological evolution and evolutionary computing sought to explain the origin of information through some process that directly used or else mimicked natural selection. Yet rather than admit a fundamental gap in explanation, this literature simply invoked selection as a backstop to explain the origin of information, the backstop itself being exempt from further explanation.

The move to explain the origin of information by invoking some separate unexplained source of information, typically via a selection process, was so common in the evolutionary literature that it deserved its own name: *displacement*.[15] *Displacement*, in general, may be defined as explaining one item of information by invoking another unexplained item of information, thereby leaving the original item of information unexplained. For instance, Alice submits a brilliant essay for her English composition class even though she has never shown any talent for writing. To explain her success, she says that Bob helped her with the essay. But Bob likewise has shown no talent for writing. In this way, Alice displaces credit that would otherwise go to her to Bob, but because Bob's writing abilities are themselves in doubt, the brilliance of her essay remains unexplained.

Displacement became the tool of choice among evolutionary critics of intelligent design as they tried to invalidate the logic of the design inference, which inferred design for events both specified and improbable (therefore exhibiting specified complexity—see the previous section). Critics claimed that once natural selection came into play, it acted as a probability amplifier that removed any seeming improbability that might otherwise have made for a valid design inference. Accordingly, critics argued that seeming products of design could be explained away through evolutionary processes requiring no design.[16]

But this attempt to invalidate the design inference was too easy. Products can be designed, but also processes that build products can be designed (compare a Tesla automobile with a Tesla factory that builds Tesla automobiles—both are designed). The design inference makes sense of improbable products. Conservation of information, through the search for a search, makes sense of improbable processes that output probable products. Making sense of displacement was a crucial step in developing a precise mathematical treatment of conservation of information.

Whereas conservation of information was a mathematically proven theoretical finding, displacement was an inductively confirmed empirical finding. Over and over, information supposedly created from scratch was surreptitiously introduced under the pretense that the information was already adequately explained when in fact it was merely presupposed. In effect, displacement became a special case of the fallacy of begging the question, obscuring rather than illuminating evolutionary processes.

One of the more brazen examples of the displacement fallacy that I personally encountered occurred in a 2001 interview with Darwinist Eugenie Scott on Peter Robinson's program *Uncommon Knowledge*. Scott and I were discussing evolution and intelligent design when Robinson raised the trope about a monkey, given enough time, producing the works of Shakespeare by randomly typing at a typewriter. Scott responded by saying that contrary to this example, where the monkey's typing merely produces random variation, natural selection is like a technician who stands behind the monkey and whites out every mistake the monkey makes in typing Shakespeare.[17] But where exactly do you find a technician who knows enough about the works of Shakespeare to white out mistakes in the typing of Shakespeare? What are the qualifications of this technician? How does the technician know what to erase? Scott never said. That is displacement: The monkey's success at typing Shakespeare is explained, but at the cost of leaving the technician who corrects the monkey's typing unexplained.

In his book *The Blind Watchmaker*, Richard Dawkins claims to show how natural selection can create information via his well-known METHINKS IT IS LIKE A WEASEL computer simulation [21, pp. 45–50]. Pure random sampling of the possibilities for a 28-character sequence consisting of letters or spaces would have a probability of only 1 in 27^{28}, or roughly 1 in 10^{40}, of achieving the target. In evolving METHINKS IT IS LIKE A WEASEL, Dawkins' simulation was able to overcome this improbability by carefully choosing a fitness landscape to assign higher fitness to character sequences that have more corresponding letters in common with this target phrase.

Essentially, in place of pure randomness, Dawkins substituted a hill-climbing algorithm with exactly one peak and with a clear way to improve fitness at any place

[15] My first serious treatment of displacement occurred in chapter 4 of [18].

[16] For an account of natural selection as a probability amplifier as well as a refutation of its use to overturn the logic of the design inference, see [19, ch. 7].

[17] "Darwinism under the Microscope," PBS television interview of William Dembski and Eugenie Scott by Peter Robinson for *Uncommon Knowledge*, filmed December 7, 2001, on the Stanford campus, with video available online at https://www.hoover.org/research/darwin-under-microscope-questioning-darwinism (last accessed December 9, 2024).

away from the peak (smooth and increasing gradients all the way!).[18] But where did this fitness landscape come from? Such a fitness landscape exists for any possible target phrase whatsoever, and not just for METHINKS IT IS LIKE A WEASEL. Dawkins explains the evolution of METHINKS IT IS LIKE A WEASEL in terms of a fitness landscape that with high probability allows for the evolution to this target phrase. Yet he leaves the fitness landscape itself unexplained. In so doing, he commits a displacement fallacy.[19]

Displacement is also evident as Dawkins moves from computer simulations to biological evolution. Indeed, his entire book *Climbing Mount Improbable* can be viewed as an exercise in displacement as applied to biology [23]. In that book, Dawkins compares the emergence of biological complexity to climbing a mountain. He calls it Mount Improbable because if you had to get all the way to the top in one fell swoop (that is, achieve a massive increase in biological complexity all at once), it would be highly improbable. But does Mount Improbable have to be scaled in one leap? Darwin's theory purports to show how Mount Improbable can be scaled in small incremental steps. Thus, according to Dawkins, Mount Improbable always has a gradual serpentine path leading to the top that can be traversed in baby steps.

But where is the verification for this claim? It could be that Mount Improbable is sheer on all sides and getting to the top via baby steps is effectively impossible. Consequently, it is not enough to presuppose that a fitness-increasing sequence of baby steps always connects biological systems. Such a connection must be demonstrated, and to date it has not, as Michael Behe's work on irreducible complexity shows [20]. But even if such a connection between biological systems linked via baby steps could be demonstrated, what would this say about the conditions for the formation of Mount Improbable in the first place—the topography over which a gradual form of evolution is said to take place?

Mountains, after all, do not magically materialize—they have to be formed by some process of mountain formation. Of all the different ways Mount Improbable might have emerged, how many are sheer so that no gradual path to the summit exists? And how many do allow a gradual path to the summit? A Mount Improbable with gradual paths to the top might itself be improbable. Dawkins simply assumes that Mount Improbable must be such as to facilitate Darwinian evolution. But in so doing, he commits a displacement fallacy, presupposing what must be explained and justified and illicitly turning a problem into its own solution.[20]

Mount Improbable is, of course, not to be taken literally. It is a metaphor for the evolvability of living things. In general, evolvability denotes the degree to which evolution is possible among different members of a configuration space in relation to the types of evolutionary change permitted (e.g., small-scale local changes versus large-scale saltational changes). Evolvability thus asks how much evolution is possible given a particular form of evolution over a configuration space. Like most Darwinists, Dawkins assumes, without adequate evidence, that a natural-selection-based form of evolution can bring about all the evolutionary connectedness required by Darwinian theory. This theory requires a lot of evolvability. Yet the evidence to justify the level of evolvability that the theory needs is not there. I return to this point in an upcoming section on evolvability (Section 11).

In the evolutionary computing literature, examples of displacement more sophisticated than Dawkins' WEASEL can readily be found. But the same question-begging displacement fallacy underlies all these examples. The most widely publicized instance of displacement in the evolutionary computing literature appeared in *Nature* back in 2003. Richard Lenski, Charles Ofria, Robert Pennock, and Christoph Adami had developed a computer simulation called Avida [24]. They claimed that this simulation was able to create complex Boolean operators without any special input or knowledge. One of the coauthors, Pennock, then went further to claim that Avida decisively refuted Michael Behe's work on irreducible complexity.[21] And given that irreducible complexity is a linchpin of intelligent design, Pennock in effect claimed that Avida had also refuted intelligent design.

But in fact, as Winston Ewert and George Montañez showed by tracking the information flow through Avida, the amount of information outputted through newly formed complex Boolean operators never exceeded the amount of prior information inputted. In fact, Avida was jury-rigged to produce the very complexity it was claiming to produce for free: Avida rewarded ever-increasing complexity simply for complexity's sake and not for independent functional reasons. Other examples like Thomas Schneider's ev, Thomas Ray's Tierra, and David Thomas's Steiner tree search algorithm all followed the

[18] For hill climbing, see [22].

[19] For a counter-simulation of the Dawkins WEASEL simulation, see "Weasel Ware – Evolutionary Simulation" by Winston Ewert and George Montañez at https://www.evoinfo.org/weasel.html. This counter-simulation shows how sensitive Dawkins' simulation is to initial inputs and how easily it is set adrift when the fitness landscape is not as neat and tidy as Dawkins' simulation demands.

[20] The three previous paragraphs are drawn in part from a lecture I gave at Oxford University's Ian Ramsey Centre on October 30, 2003 titled "Gauging Intelligent Design's Success." Though on faculty at Oxford, Richard Dawkins was not in attendance. The lecture is available at https://billdembski.com/documents/2003.11.Gauging_IDs_Success.pdf (last accessed December 13, 2024).

[21] Pennock, citing the 2003 *Nature* article, claimed that "colleagues and I have experimentally demonstrated the evolution of an IC system." IC here is "irreducibly complex." Quoted from [25, p. 141].

same pattern.[22] Ewert and Montañez were able to show precisely where the information supposedly created from scratch in these algorithms had in fact been embedded from the outset.[23] Displacement, as their research showed, is pervasive in this literature.

The empirical work of showing displacement for these computer simulations set the stage for the theoretical work on conservation of information. These simulations, and their consistent failure to explain the origin of information, prompted an investigation into the precise numerical relation between information inputted and information outputted. Showing displacement started out as a case-by-case effort to uncover where precisely information had been smuggled into a computer simulation. Once the mathematics of conservation of information was developed, however, the need to find exactly where the information was smuggled in was no longer so important, theory stepping in where observation fell short.

Theory guaranteed that the information was smuggled in even as the evolutionary simulations became so byzantine that it was hard to follow their precise information flow. By analogy, if you have a hundred and one letters that must go into a hundred mail boxes, the pigeonhole principle of mathematics guarantees that at least one of the mailboxes must have more than one letter [28, p. 10]. Checking this empirically could be arduous if not practically impossible because of all the many possible ways that these letters could fill the mailboxes. Theory in this case comes to the rescue, guaranteeing what observation alone cannot.

Displacement is a shell game. In a shell game, an operator places a small object, like a pea, under one of three cups and then rapidly shuffles the cups to confuse observers about the object's location. Participants are invited to guess which cup hides the pea, but the game often relies on sleight of hand and misdirection to increase the likelihood that participants guess incorrectly. So long as the game is played fairly, the pea is under one cup and remains under that same cup.[24] It cannot magically materialize or dematerialize. The game can become more sophisticated by increasing the number of cups and by the operator moving the cups with ever greater speed and agility. But by carefully tracking the operator, it is always possible to determine where the pea started out and where it ended up. The pea here is information. Displacement says that it was always there. Conservation of information provides the underlying mathematics to demonstrate that it was indeed always there.

9. A BAYESIAN PROOF OF THE LAW OF CONSERVATION OF INFORMATION

A probability measure P on a probability space Ω assigns a probability to a certain class of subsets of Ω according to the axioms of probability. If Ω is countable, then P is defined on all subsets. For bigger Ω, P is defined on the measurable subsets, often called Borel sets. When Ω is a search space and P gauges the probability of successful search, the focus is on the probability of targets T in the search space Ω. However, that search space does not incorporate the search-for-a-search, or S4S, space that is central to conservation of information. So, the question arises where to fit this S4S space in relation to the original search space.

To see what is at issue here, recall the die-tossing example with which we started this paper. The search space Ω consisted of six possible die rolls—namely, {1, 2, 3, 4, 5, 6}. The corresponding S4S space, however, consisted of all loadings of the die that for any two faces assigned a probability of $\frac{1}{2}$ to either face. That higher-level S4S space, which in Section 3 of this paper was denoted by $\overline{\Omega}$, consisted of fifteen possible loadings of pairs of faces, and was represented as {(1,2), (1,3), (1,4), (1,5), (1,6), (2,3), (2,4), (2,5), (2,6), (3,4), (3,5), (3,6), (4,5), (4,6), (5,6)}. In this representation, each ordered pair denotes a loading of the corresponding pair of faces, where each face in the pair has the same probability of $\frac{1}{2}$ of landing face up.

Ω and $\overline{\Omega}$ are therefore quite different search spaces, and so the challenge is to coordinate them when calculating the probabilities relevant to conservation of information. Thus, in this die-loading example, where the aim is to roll a six, the target T is {6}, and its probability p is equal to $\frac{1}{6}$. But if the aim is to roll a six with one of the loaded dice in $\overline{\Omega}$, then the higher-level target is $\overline{T}$ = {(1,6), (2,6), (3,6), (4,6), (5,6)}, whose probability in $\overline{\Omega}$ is $r = \frac{1}{3}$. Moreover, each element of this higher-level target is interpreted as a search that locates the original target T with probability $q = \frac{1}{2}$. Thus, it is natural to want to write $P(T) = p$, $P(T|\overline{T}) = q$, and $P(\overline{T}) = r$ to refer to these probabilities. And in fact, we will see shortly that it is mathematically legitimate to write these probabilities in this way. Yet for now, with the probability measure P only defined on Ω, such

[22] For ev, see [26]. For the best place to understand Tierra, see Thomas Ray's website https://tomray.me/tierra. For a search algorithm purported to solve the Steiner Tree problem without the need for full prior information, see [27] and then a follow-up by Thomas titled "Target? TARGET? We do not Need No Stinkin' Target!" https://pandasthumb.org/archives/2006/07/target-target-w-1.html (last accessed December 10, 2024).

[23] See the counter-simulations by Ewert and Montañez at EvoInfo.org: contra Avida, see their "Minivida—Dissection of Avida Digital Evolution" at https://www.evoinfo.org/minivida; contra ev, see their "Ev Ware – Evolutionary Simulation" at https://www.evoinfo.org/ev (last accessed December 13, 2024). See also [2], where we critique all these evolutionary simulations that purport to create novel information that exceeds their prior informational input. Dave Thomas is critiqued in this book on pages 119–120 and 241–242.

[24] Often it is played unfairly, as when the pea is surreptitiously removed from under one of the shells. For how this is done, see https://www.youtube.com/watch?v=GumWeVdcof4 (last accessed February 21, 2025).

an assignment of probabilities is not, strictly speaking, allowed. To allow it, we need to extend the probability measure P to encompass the original search space Ω and the S4S space $\overline{\Omega}$. But how to do that?

Fortunately, we have a model from Bayesian theory of how to extend a probability P from the original search space Ω to encompass also the S4S space $\overline{\Omega}$. The Bayesian approach to probability coordinates probabilities between events and hypotheses. Thus, as with search, we have a probability P that is defined on a space Ω of events, which in search correspond to targets. However, when we introduce a collection of hypotheses, as in Bayes' theorem, the probability space is effectively expanded beyond just Ω. Thus, instead of considering a single probability measure, we must now work with a family of probability measures $P(\cdot|H)$, indexed by hypotheses H in some space of hypotheses $\mathcal{H}$ ($\mathcal{H}$ corresponds to $\overline{\Omega}$).

The Bayesian framework focuses on such conditional probabilities to quantify changes in uncertainty in the face of new evidence. Managing uncertainty within the Bayesian framework also requires assigning prior probabilities $P(H)$ to the hypotheses H in $\mathcal{H}$. Accordingly, a probability measure on the original probability space Ω needs, within the Bayesian framework, to be extended to a probability measure on the Cartesian product space $\Omega \times \mathcal{H}$ (or equivalently on the Cartesian product space $\Omega \times \overline{\Omega}$, since we are identifying $\mathcal{H}$ and $\overline{\Omega}$).

Within the Bayesian framework, probabilities of events conditional on hypotheses are called *likelihoods.* A likelihood of the form $P(E|H)$ represents the probability of observing an event E given some specific hypothesis H. When the hypothesis space $\mathcal{H}$ is discrete (finite or countably infinite), Bayes' theorem calculates a probability based on a weighted average over all possible hypotheses, as appears in the denominator of this standard formulation of Bayes' theorem:

$$P(H|E) = \frac{P(E|H)P(H)}{\sum_{H' \in \mathcal{H}} P(E|H')P(H')}.$$

The summation in this denominator accounts for all the ways E could occur under the different hypotheses in $\mathcal{H}$, weighted by their prior probabilities. It equals $P(E)$ because the hypotheses in $\mathcal{H}$ are assumed to be mutually exclusive and exhaustive.

When the hypothesis space $\mathcal{H}$ is continuous (noncountable), the summation in this denominator is replaced by integration over a probability measure, which can be an arbitrary probability measure but typically takes the form of a probability density function $f(H)$ integrated with respect to a canonical measure dH (typically Lebesgue measure over a Euclidean space). The probability of E in the denominator of Bayes' theorem is then obtained by integrating the likelihood over all hypotheses:

$$P(E) = \int_{\mathcal{H}} P(E|H)f(H)dH.$$

Bayes' theorem then updates the hypothesis density distribution as follows:

$$f(H|E) = \frac{P(E|H)f(H)}{\int_{\mathcal{H}} P(E|H')f(H')dH'}.$$

In the continuous case, an integral averages out the likelihood over the hypothesis space, whereas in the discrete case a summation does the work of averaging. Thus, in a discrete hypothesis space, we compute posterior probabilities (i.e., probabilities of the form $P(H|E)$) by summing contributions from each discrete hypothesis, while in a continuous hypothesis space, we weight all possible hypotheses through integration. This distinction proves important in Bayesian inference, as many real-world problems involve continuous parameter spaces (e.g., estimating a population mean or variance), requiring the use of probability densities rather than finite discrete assignments of probabilities.[25] Nonetheless, for ease of exposition, we will focus in this paper entirely on the finite discrete case. Many S4S spaces are in fact spaces of probability measures, and such spaces invariably fall under the continuous case. But the continuous case is a straightforward generalization of the discrete case, so in effect focusing on the discrete case involves no loss of generality.[26]

The Bayesian framework just outlined can be repurposed to coordinate search with search for a search. We can think of a target T as an event E that locates the target. The probabilities of T and E are then equivalent. Moreover, we can think of a higher-level target $\overline{T}$ of searches that find targets in the original search space as a collection of hypotheses each of whose elements is a hypothesis H specifying a search of the original search space. Each such search then induces a probability measure corresponding to the search's probability of finding a target, or, equivalently in our repurposed framework, a hypothesis that characterizes a search that locates a target. Targets of the original search thus become events,

[25] For Bayesian mathematics in the continuous case, see [29, pp. 4–5].

[26] For a finite probability space of two or more elements, the space of probabilities atop it consists of all convex linear combinations of elements from the original space. So for an underlying space of $n \geq 2$ elements, the space of probabilities on those n elements is a simplex of dimension n–1 in an n-dimensional Euclidean space. For infinite underlying probability spaces with some minimal structure (i.e., embeddable into a complete separable metric space), the space of probability measures over them consist of weak limits (in the sense of weak convergence) of convex linear combinations of finitely many discrete point masses from the original space. The underlying measure theory here is more than most readers of this paper will likely desire. For the details, see [30].

and higher-level targets for locating lower-level targets thus become collections of hypotheses.

Within the Bayesian framework, a probability measure P therefore exists on the joint Cartesian product space $\Omega \times \mathcal{H}$, where Ω denotes the original search space, conceived in terms of events that locate targets, and $\mathcal{H}$ denotes the S4S space, conceived in terms of hypotheses that characterize searches of the original search space. For an event E, $P(E)$ denotes the marginal probability distribution on the first factor of this product and is therefore equated with $P(E \times \mathcal{H})$. Similarly, for a collection of hypotheses F that correspond to one or more searchs (namely, $F = \{H_i\}_{i \epsilon J} \subset \mathcal{H}$ for some indexing set J), $P(F)$ denotes the marginal probability distribution on the second factor of this product, and is therefore equated with $P(\Omega \times F)$. The joint probability of E and F—namely, $P(E \cap F)$—is then defined as:

$$P((E \times \mathcal{H}) \cap (\Omega \times F)),$$

and the conditional probability $P(E|F)$ is then defined as:

$$P((E \times \mathcal{H})|(\Omega \times F)) = \frac{P((E \times \mathcal{H}) \cap (\Omega \times F))}{P(\Omega \times F)}.$$

It is conventional to use a minimalist notation in Bayesian probability analysis. Thus, for marginal probability we write $P(E)$ rather than $P(E \times \mathcal{H})$ and $P(F)$ rather than $P(\Omega \times F)$. For conditional probability, we write $P(E|F)$ rather than the more complicated and formally explicit $P((E \times \mathcal{H})|(\Omega \times F))$. Because it is typically clear what is happening on the event side and on the hypothesis side in a Bayesian analysis, this "abuse of notation" is innocent, and one might even say virtuous, by keeping the notation clean and concise.

Some idiosyncrasies arise when the Bayesian formalism is applied to search and consequently to the proof of the Law of Conservation of Information. In many situations where Bayes' theorem is applied, the hypotheses in question are few. A Bayesian analysis, for instance, might assess to what degree a positive result on a medical test establishes that the person taking the test actually has the disease. The hypotheses in question then are "the test gave an accurate reading" (correct positive) versus "the test gave a false reading" (false positive). Together these two hypotheses are mutually exclusive and exhaustive. Moreover, based on past medical data, their prior probabilities would be known with great accuracy.

By contrast, in search, the hypothesis space $\mathcal{H}$, which corresponds to the search-for-a-search or S4S space, will typically be much larger than the original search space. Consider again our die-tossing example in which the aim was to roll a six. The search space consisted of six items (the different faces of the die). But the S4S space consisted of fifteen items—namely, all the ways the die could be loaded so that any two given faces have probability $\frac{1}{2}$ of turning up. But in fact, the hypothesis space, or S4S space, here could have been much larger.

What if we had allowed any loading of the die whatsoever? In that case, the S4S space would consist of all 6-tuples of weights $(w_1, w_2, w_3, w_4, w_5, w_6)$ where each weight is non-negative and together they satisfy $w_1 + w_2 + w_3 + w_4 + w_5 + w_6 = 1$. In other words, the S4S space would now be a 5-dimensional simplex residing in a 6-dimensional Euclidean space. Its size would therefore be not just infinite but have the cardinality of the continuum (i.e., the order of infinity of the real numbers). Thus, a puny search space of six elements here gives rise to a huge S4S space.[27]

Although Bayesian theory does not care about the size of the hypothesis space $\mathcal{H}$, it tends to focus on individual hypotheses one at a time. Thus, in a Bayesian analysis, probabilities involving hypotheses usually take the form of $P(E \cap H)$ or $P(E|H)$, where H is a single hypothesis in $\mathcal{H}$. But there is nothing in the Bayesian formalism to prevent probabilities from taking the form of $P(E \cap F)$ or $P(E|F)$, where $F = \{H_i\}_{i \epsilon J}$ is a collection of hypotheses/searches in $\mathcal{H}$. The rules of probability apply and everything comes out in the wash. Moreover, if $\mathcal{H}$ is a continuous space, such as having the structure of an n-dimensional simplex, then a continuous form of the Bayesian formalism applies, and again everything comes out in the wash. As said, we focus on the discrete case for simplicity but also because it entails no substantive loss of generality.

In a typical application of Bayesian theory, the hypothesis space $\mathcal{H}$ can, as noted, be quite small, as in a medical test that either does or does not give a true reading and so involves only two hypotheses. Conversely, as also noted, in the context of search, the hypothesis space $\mathcal{H}$, representing the S4S space, is typically quite large and at least as large as the original search space. The precise degree of largeness of the S4S space can vary. However, the S4S space needs to be large enough in the sense that it satisfies what may be called a *full-spectrum search condition.* What this means is that the S4S space must be adequate for representing the full range of relevant searches.

If the target T is really hard to find (i.e., it is a needle in a haystack), $P(E)$, the probability of the event that corresponds to finding the target, is going to be very small. But what does that say about the hypotheses *qua* searches that make up the hypothesis or S4S space $\mathcal{H}$ (also denoted $\overline{\Omega}$)? Because $P(E) = \sum_{H \in \mathcal{H}} P(E|H) \times P(H)$, the hypotheses/searches in $\mathcal{H}$ must, on average, confer a small probability on E. At the same time, we are, in the context of conservation of information, concerned with identifying searches that with high probability locate the target. This means that there will be some hypotheses $H \in \mathcal{H}$

[27] See the previous note.

such that $P(E|H)$ will be large. But these large probabilities will then need to be down-weighted by small probabilities $P(H)$ since otherwise $P(E)$ will itself become large, contrary to assumption. Simply put, if $P(E|H) \gg P(E)$ then $P(H) \ll 1$.

To illustrate the full-spectrum search condition, consider the earlier example of an Easter egg hunt. An egg can be hidden anywhere on a large field, and so the search space must be capable of representing all these places the egg could be hidden. So too, instructions for finding the egg must be able to guide the Easter egg hunter to any place on the field where an egg might be hidden. The collection of all such instructions would then constitute the corresponding S4S space. In this example, the S4S space therefore satisfies the full-spectrum search condition.

On the other hand, a violation of the full-spectrum search condition would occur if places where the egg could be hidden were arbitrarily restricted or certain instructions for finding eggs were arbitrarily ruled out. Such violations would introduce bias into the S4S space, possibly making the egg more probable or less probable to be found than its actual probability. In general, we should be suspicious of search spaces that fail to represent the full range of search possibilities (such as omitting a portion of the field where the egg might be hidden, in our example). Likewise, we should be suspicious of S4S spaces that fail to represent the full range of ways the original search space can be searched. Simply put, we should be suspicious whenever the full-spectrum search condition is violated.

Yet ultimately, for the Law of Conservation of Information to hold, it must be the case that:

$$P(E) = P(E \times \mathcal{H})$$

or equivalently:

$$P(E) = \sum_{H \in \mathcal{H}} P(E|H) \times P(H).$$

In particular, the probability of E, when conditioned on and averaged across all searches in the S4S space, will need to be small. In practice, a typical search space permits searching for many different targets distributed widely across the search space. Consequently, its corresponding S4S space will need to include enough searches to be able to search for all these different targets, and that includes varying the degree of probability of these searches in finding the different targets. In other words, the S4S space needs to be adequate for representing all the different searches that might arise on the original search space. Or, as we might say, given the full-spectrum search condition, the S4S space needs to exhibit a full-spectrum of searches.

The Law of Conservation of Information depends critically on conditioning not just on individual hypotheses or searches but on entire collections of them simultaneously. Given an event E (i.e., an event of finding a target) and a single hypothesis H (i.e., a hypothesis characterizing a single search) where the baseline probability is p and the improved probability is q, it follows that $P(E) = p$ and $P(E|H) = q$. If we now let $P(H) = r$, it follows as a matter of elementary probability that $P(E) \geq P(E \cap H) = P(E|H) \times P(H)$, and so $p \geq q \times r$, or $\frac{p}{q} \geq r$, which as we saw in Section 5 is the defining inequality for the Law of Conservation of Information.

However, the full Law of Conservation of Information requires the ability to handle entire collections of hypotheses at once. Suppose, therefore, that we are given an event E that denotes finding a target T in Ω. And suppose further that we are given a collection of hypotheses $F = \{H_i\}_{i \epsilon J}$ each of which hypothesizes a search in Ω, assigning a probability for any possible target. Each of the hypothesized searches is presumed to be distinct and to be mutually exclusive of the others. By substituting $\{H_i\}_{i \epsilon J}$ for F, it then follows from elementary probability theory that:

$$\begin{aligned} P(E) &\geq P(E \cap F) \\ &= P(E \cap \{H_i\}_{i \in J}) \\ &= \sum_{i \in J} P(E \cap H_i) \\ &= \sum_{i \in J} P(E|H_i) \times P(H_i). \end{aligned}$$

Suppose now, as is typical with conservation of information, that $P(E) = p$, where p is a very small. $P(E)$ is then a *baseline probability*. Suppose further that F denotes a collection of hypotheses that jointly correspond to a higher-level S4S where each search in it elevates the probability of finding E to a probability q or greater, q being much bigger than p. F therefore corresponds to a higher-level target where for each i in the index set J, $P(E|H_i) \geq q$. This probability q is therefore the *improved probability*. Consequently, $P(F) = r$ is the *S4S probability*. It then follows that:

$$\begin{aligned} p &= P(E) \\ &\geq \sum_{i \in J} P(E|H_i) \times P(H_i) \\ &\geq \sum_{i \in J} q \times P(H_i) \\ &= q \times P(F). \end{aligned}$$

The second summation equals $q \times P(F)$ because the probability q pulls out of the summation and the summation of the $P(H_i)$ over the index set J is $P(\{H_i\}_{i \epsilon J})$ because the hypotheses/searches are distinct/mutually exclusive and because $\{H_i\}_{i \epsilon J}$ is F. From this inequality, it now follows that $\frac{p}{q} \geq P(F) = r$. And so, the baseline probability divided by the improved probability is greater than or equal to the S4S probability. Substituting T for E and $\overline{T}$ for F, this inequality matches up with conservation of information as characterized mathematically in Section 5.

We are close to the full statement and proof of the Law of Conservation of Information, but we are not quite there yet. The remaining point at issue is the role of $F = \{H_i\}_{i\epsilon J}$ in the following inequality:

$$P(E) \geq P(E \cap F) = P(E|F) \times P(F).$$

This inequality holds as a matter of elementary probability theory. If $F = \{H_i\}_{i\epsilon J}$ consists of multiple hypotheses (searches) and if $P(E|F) = q$, what does this say about the effectiveness of the original searches/hypotheses H_i in F for finding the original target, this target being represented by the event E? As it is:

$$\begin{aligned} q &= P(E|F) \\ &= P(E|\{H_i\}_{i\in J}) \\ &= \sum_{i\in J} P(E|H_i) \times \frac{P(H_i)}{P(F)}. \end{aligned}$$

But because $F = \{H_i\}_{i\epsilon J}$, it follows that the weights $\frac{P(H_i)}{P(F)}$ when summed over the index set J equal 1. Consequently, the conditional probabilities $P(E|H_i)$, when averaged by these weights, equals q. The individual conditional probabilities $P(E|H_i)$ must therefore either be at least q or, if falling below q, must be counterbalanced by other conditional probabilities for hypotheses in $F = \{H_i\}_{i\epsilon J}$ that are greater than q. In other words, average performance of searches in F will equal q when $P(E|F)$ equals q. This is what we need for a general statement and proof of the Law of Conservation of Information, making clear what it means for $P(E|F)$ to equal q.

To tie it all together, given that $P(E) \geq P(E|F) \times P(F)$, with $P(E)$ being very small (equal to p) and $P(E|F)$ large (equal to q, often close to or even equal to 1), it follows that $\frac{p}{q} \geq P(F)$. If p is close to zero and q is close to one, it follows that $\frac{p}{q}$ will itself be close to zero and so $P(F)$, the probability of a successful search for a search, will itself be very small. What the inequality $P(E) \geq P(E|F) \times P(F)$ means for search now becomes clear. If $P(E)$ is very small because E by itself, and thus without additional aid, is a needle-in-a-haystack problem, and if $P(E|F)$ is large because a search for search (S4S) has successfully found a collection of searches F that, on average, with high probability bring about E, then the probability of F left to itself must itself be small. Simply put, $P(F)$ must be small to offset $P(E|F)$ being large so that the product $P(E|F) \times P(F)$ is small enough to be less than or equal to $P(E)$, which is known to be small. Put even more simply, insofar as F elevates the probability of E, F must depress its own probability.

What may legitimately be called the Law of Conservation of Information now immediately follows and is readily stated:

> *Given a search having a probability of success p for which a search for a search raises the probability of success to q, the probability of successfully achieving such a search for a search will be less than or equal to* $\frac{p}{q}$.

As formulated within a Bayesian framework, the proof of the Law of Conservation of Information is immediate from the inequality $P(E) \geq P(E|F) \times P(F)$. Therefore, this law is a direct consequence of probability theory as applied to search. It is not an inductive generalization, as might arise by enumerating positive instances of some phenomenon. Nor is it a logical consequence of an empirically based theory, such as the laws of thermodynamics, where theory and observation mix. It is a simple fact of pure mathematics.

10. CLARIFYING THE LAW OF CONSERVATION OF INFORMATION

The Bayesian proof of the Law of Conservation of Information given in the last section is technically sound. This section helps to clarify its statement and proof, especially in how to interpret and apply conservation of information in practice.

10.1 Convenience for General Use

Some heavy-duty mathematics is implicit in the Law of Conservation of Information, especially if we try to unpack the mathematical structure of the S4S spaces that arise in practice, which often are measure spaces with convergence properties whose full understanding requires such advanced mathematics as measure theory and functional analysis (see Appendix 2). Moreover, even the full Bayesian formalism into which we cast the Law of Conservation of Information can be more than many users of the law might feel technically equipped to handle. As it is, for most practical purposes, it is enough simply to focus on the inequality $P(E) \geq P(E|F) \times P(F)$, which mutes the underlying Bayesian formalism and rationale while also bypassing any still more demanding mathematics.

The inequality $P(E) \geq P(E|F) \times P(F)$ is convenient for general use, respecting our intuitions about the law and allowing for reliable applications of it. Take, for instance, an example given by Douglas Axe.[28] He considers the event E of an arrow hitting a bullseye. If the arrow is shot completely at random (without the archer

[28] I got this example from Axe's talk at the intelligent design symposium of the Evangelical Philosophical Society meeting in San Diego, November 21, 2024.

even taking aim), the probability $P(E)$ will be extremely small—call it p. But imagine now a wind that blows the arrow so it hits the bullseye with high probability. Call the event of such a wind blowing F. Because the wind blows the arrow to favor its landing on the target, $P(E|F)$ is now large—call it q, where q is much larger than p ($q \gg p$). But because $P(E) \geq P(E|F) \times P(F)$, it then follows that $\frac{p}{q} \geq P(F)$, implying that getting the right wind so that the arrow lands on the target offers no advantage over hitting the target directly by random chance.

This intuitive approach to setting up search problems involving small probabilities and then reasoning about them falls readily under conservation of information. Yet for practical purposes, nothing is lost by omitting explicit mention of the Bayesian formalism or any deeper mathematics. To analyze the problem and then reach the right conclusion, it is generally enough, as in the previous example, to work directly with the inequality $P(E) \geq P(E|F) \times P(F)$, which we may dub the *conservation of information inequality.*

10.2 No Subverting the Law Incrementally

Applying the Law of Conservation of Information becomes interesting when the event in need of explanation (hitting the target) has a small probability. One approach to explaining away small-probability events is then to reason that the probabilities really are not all that small once all the opportunities for the event to occur are taken into account. Consider, for instance, a bullet hitting a fly on a large wall. The event of that bullet hitting that particular fly might be vastly improbable. But what if the probability of the bullet hitting the wall was quite large and what if the wall was plastered with flies? In that case, hitting some fly with even a single bullet could become quite likely. Incremental ways like this of adding in opportunities to circumvent small probabilities have a long history. But this approach does not always work and, as we will see next, it does not refute the Law of Conservation of Information.

In the context of search, the concern just raised can be stated as follows: Can multiple hypotheses (or searches) collectively raise the probability of finding a target to a level that circumvents the Law of Conservation of Information? Are there effective ways to combine searches so as to attain E (i.e., locate the target) without paying the full probabilistic cost required by the law? Through some clever probabilistic manipulation, is it possible to sidestep the Law of Conservation of Information and thereby obtain something like a free lunch?

Formulated for individual hypotheses, the Law asserts that $P(E) \geq P(E|H) \times P(H)$. The key probability in question here is the S4S probability, which in Bayesian terms is the prior probability $P(H)$. This probability tends always to be the pivotal one in any Bayesian analysis as it reflects initial assumptions before considering evidence, which means that any Bayesian reasoning is always downstream of it. It is also the probability most in question in any Bayesian recasting of search and in particular with the search for a search. The key claim of the Law of Conservation of Information is that the S4S probability $P(H)$ is always so small that it cannot, when all is said and done, raise the probability of successful search above the probability of the original search (i.e., $P(E)$).

So how, if at all, does the Law of Conservation of Information need to be adjusted when the search for a search yields not just a single search but a collection of searches? In other words, what happens when the search for a search delivers not a single H but a collection of Hs, such as $F = \{H_i\}_{i\epsilon J}$? In fact, all the key probabilistic relationships specified by the Law of Conservation of Information continue to hold. The S4S probability is then $P(F)$ for a collection of searches. From elementary probability theory it follows that $P(E) \geq P(E|F) \times P(F)$. But if the baseline probability is $p = P(E)$ and the S4S probability is $P(F)$, it does not follow that if every search (hypothesis) H_i in $F = \{H_i\}_{i \in J}$ raises the probability of E's occurrence (i.e., the finding of the target) to a probability of q or higher that $P(E|F)$ is actually equal to q. In fact, $P(E|F)$ will be at least as large as q.

To see this, rewrite:

$$P(E|F) \times P(F) = P(E|\{H_i\}_{i \in J}) \times P(\{H_i\}_{i \in J})$$

as the following:

$$P(E|F) \times P(F) = \sum_{i \in J} P(E|H_i) \times P(H_i)$$

Then, because each $P(E|H_i)$ is greater than or equal to q, it follows that this sum in turn is greater than or equal to $P(F) \times q$, from which the crucial inequality $\frac{p}{q} \geq P(F)$ then immediately follows. The S4S probability is therefore still small, even with this maneuver of incrementally adding hypotheses.

But what if we do not just cumulate individual hypotheses, but sets of hypotheses? Suppose, for instance that $F = F_1 \cup F_2 \cup \cdots \cup F_n$ for some n greater than 1, and where each of the terms in this union itself consists of individual hypotheses. To avoid overcounting, we assume each of the F_i's do not intersect and so are mutually exclusive. Each of the F_i's for $1 \leq i \leq n$ will now, as a matter of basic probability theory, satisfy $P(E) \geq P(E|F_i) \times P(F_i)$. But if each of F_i raises the improved probability $P(E|F_i)$ to q or more, which is to say $P(E|F_i) \geq q$, then, because the F_i's are mutually exclusive:

$$\begin{aligned} P(E) &\geq P(E|F) \times P(F) \\ &= \sum_{1 \leq i \leq n} P(E|F_i) \times P(F_i) \\ &\geq \sum_{1 \leq i \leq n} q \times P(F_i) \\ &= q \times P(F). \end{aligned}$$

Given $p = P(E)$, this again means that the S4S probability $P(F)$ remains bounded above by $\frac{p}{q}$.

Bottom line: Cumulating higher-level S4S targets each of which has probability at least q of locating the original target does nothing to improve the overall probability of finding the original target.

10.3 Conserved vs. Unconserved

Conservation of information asserts that for information to be strictly conserved is the best we can hope for with search but that often information is lost, and so we must settle for less than strict conservation of information. Let us now examine what exactly this means in the formulation of the Law of Conservation of Information. Given that $P(E) = p$ and $P(E|F) = q$, we say that information is *conserved* if $P(F)$ is strictly equal to $\frac{p}{q}$, *unconserved* if $P(F)$ is strictly less than $\frac{p}{q}$. So, when do we have strict equality and, therefore, strict conservation? And when do we have strict inequality and, therefore, loss of conservation?

If we let F^c denote the complement of F (i.e., the mutually exclusive and exhaustive alternative to F), then the union $F \cup F^c$ (i.e., the occurrence of either F or F^c) will denote the certain event, and so $P(F \cup F^c) = 1$. Because intersection distributes over union and because F and F^c together are mutually exclusive and exhaustive, it then follows from elementary probability that:

$$\begin{aligned} P(E) &= P(E \cap (F \cup F^c)) \\ &= P((E \cap F) \cup (E \cap F^c)) \\ &= P(E \cap F) + P(E \cap F^c) \\ &= [P(E|F) \times P(F)] + [P(E|F^c) \times P(F^c)]. \end{aligned}$$

In other words:[29]

$$P(E) = [P(E|F) \times P(F)] + [P(E|F^c) \times P(F^c)].$$

It follows that only if the second term on the right of this equation is zero does $P(E) = P(E|F) \times P(F)$, implying strict conservation of information. On the other hand, if $P(E|F^c) \times P(F^c)$ is non-zero, then it contributes positively to the right side of the above equation, and so $P(E) > P(E|F) \times P(F)$, implying a lack of strict conservation.

Thus, when $P(E) > P(E|F) \times P(F)$, attaining F as a way of attaining E lowers the probability of occurrence of E below the original probability $P(E)$. We saw a case in which $P(E|F^c) \times P(F^c)$ was zero, and so information was conserved in Section 2. There we considered loadings of a die that guaranteed rolling each of two faces with a probability of $\frac{1}{2}$. F in this case consisted of the five ways of loading the die so that the die face-6 had probability $\frac{1}{2}$, and F^c consisted of ten ways of loading the die so that die face-6 had probability zero. Accordingly, because $P(F^c) = \frac{2}{3}$ but $P(E|F^c) = 0$ (because all loadings in F^c render E impossible), $P(E|F^c) \times P(F^c) = 0 \times \frac{2}{3} = 0$. In Section 2 we also considered a loading where strict conservation was lost because of a slight modification of the previous example where the dominant pair of faces each had probability $(\frac{1}{2} - \varepsilon)$ for some small positive ε.

Generally speaking, strict conservation is more readily lost than retained. In practical circumstances, information is subject to a kind of friction where through use and reuse of the information, inefficiencies crop up. Perfect efficiency in information transfer is therefore an ideal that is rarely attained, a point underscored in Shannon's theory of information where noisy channels are common and undercut perfect accuracy of transmission.

10.4 Probability Diminisher vs. Probability Amplifier

Given our fundamental inequality $P(E) \geq P(E|F) \times P(F)$, from the assumption that $P(E)$ is small and $P(E|F)$ is large, it follows that the S4S probability $P(F)$ must also be small. But how do we know that the probability of E is indeed small? What if we flip the focus by starting with a large probability of F and feed it into the Law of Conservation of Information. With a large S4S probability $P(F)$ and a large improved probability $P(E|F)$, it will then follow that the $P(E)$ is itself large. Taking this changed perspective to its logical limits, it might then be argued that any needle-in-a-haystack problem that gets solved (corresponding to the occurrence of E) signifies not the occurrence of some vastly improbable event but rather the occurrence of some vastly probable prior event (corresponding to F) that allows the problem to be solved. If you will, all needle-in-a-haystack problems might thus become searches for iron needles sought by magnetic gloves that attract and extract needles from haystacks with high probability.[30]

What we have here then is a diminution-amplification dialectic: diminish E's probability and F's probability must likewise be diminished; on the other hand, amplify F's probability, and E's probability must likewise be amplified. That is assuming of course that the improved probability of E given F is large. Our emphasis in this paper has been on probability diminution: start with the

[29] For the relevant probability theory, see Appendix A of [19]. The equality $P(E) = [P(E|F) \times P(F)] + [P(E|F^c) \times P(F^c)]$ is a special case of the law of total probability, which is treated on page 416 of Dembski and Ewert's book.

[30] I am indebted to Marc Mullie for this analogy.

baseline probability $P(E)$ being small, and given a large improved probability $P(E|F)$, the S4S probability $P(F)$ will necessarily have to be small. But the emphasis could instead be put on probability amplification: start with the S4S probability $P(F)$ being large, and given a large improved probability $P(E|F)$, $P(E)$ will necessarily have to be large. Probability amplification is implicit in the evolutionary literature, where biological adaptations are viewed not as highly improbable occurrences that violate statistical norms but rather as entirely expected, and therefore probable, based on evolutionary antecedents.

Now it can happen that the event E, whose occurrence we are trying to explain, may be insufficiently understood to convincingly assign it a small probability. If so, we would need to leave open that the prior event F could have a large probability. But often we do have compelling reasons for accepting that E's probability is indeed small. Typically such reasons take the form of an argument from contingency: the event E is one of a range of events, where this range allows multiple degrees of freedom so that any event in the range is as likely to occur as any other. Consequently, if the range is vast and any event in it is highly improbable, there is no way for search processes to raise E's probability except by imposing externally applied information through some prior event F that, by the Law of Conservation of Information, must then have small probability

In practice, we establish the contingency of an event by showing that it is compatible with the mechanisms involved in its production and yet that its alternatives are equally compatible with those mechanisms. These mechanisms, typically conceived as natural laws or algorithms, are therefore necessary but insufficient to explain the event's production, failing to distinguish the event from possible alternatives. The number of these alternatives are then the degrees of freedom associated with the event. The greater these degrees of freedom, the more improbable the event. By being compatible with but not necessitated by the mechanisms involved in its production, an event cannot have anything other than a small probability. The event is radically contingent: it not only allows alternatives but also any alternatives are, given the underlying mechanisms of production, as probable as it. The event could happen but need not happen, and its happening is highly improbable.

Michael Polanyi described this method for establishing contingency via degrees of freedom in the 1960s.[31] He used this method to argue for the irreducibility of biology to physics and chemistry. The method applies quite generally:

- the position of Scrabble pieces on a Scrabble board is irreducible to the natural laws governing the motion of Scrabble pieces;
- the configuration of ink on a sheet of paper is irreducible to the physics and chemistry of applying ink to paper;
- radio signals used to convey semantic information are irreducible to the laws of physics that govern the transmission of radio signals;
- the sequencing of DNA bases is irreducible to the chemical bonding affinities between the bases (in particular, the bases are completely interchangeable along DNA's sugar-phosphate backbone);
- the sequencing of amino acids in proteins is irreducible to the peptide bonds connecting the amino acids;
- the choreography of dancers on a stage is irreducible to the biomechanics of their movements;
- the notes heard in a musical composition are irreducible to the structure and dynamics of the instruments producing the sounds;
- the positions assumed by pieces in a game of chess is irreducible on the physics governing the movement of pieces on a chessboard. Etc. Etc. Etc.

For Polanyi, such examples did not just illustrate a method for establishing contingency and therefore the improbability of events. Rather, they also illustrated how systems exhibiting higher-order information or intention transcend the lower-level physical or chemical processes that undergird them.

Radical contingency resulting from vastly many degrees of freedom does not occur in every situation, but when it does occur, it entails vast improbability of any event in the range of contingent events. Frequently, such vast degrees of freedom are associated with configuration spaces that comprise all possible sequences from a fixed and reusable set of basic elements, such as a fixed alphabet. Such spaces model written texts as well as polymers like DNA, RNA, and proteins. These configuration spaces are symmetric, meaning any elemental sequence is geometrically interchangeable with another, though not semantically. This symmetry ensures that the mechanical forces responsible for forming sequences cannot distinguish or prefer certain sequences over others. Instead, external semantic information (for written texts) or functional information (for biopolymers) is required to differentiate sequences constructed from a fixed set of basic elements, such as an alphabet. Claiming that this information reduces to material mechanisms is like arguing that Scrabble pieces inherently prefer specific arrangements, which they clearly do not.[32]

[31] See [31]; For the same line of reasoning, see also [32, pp. 7–8] as well as [33, p. 335].

[32] For further exposition of this argument from contingency for DNA, see [34].

There is also a practical reason to think that in applying the Law of Conservation of Information, any probability we assign to E (the baseline search) will on the whole be more credible than any probability we assign to F (the S4S search). E will typically be an event that confronts us. To engage our interest, it needs to have actually occurred (or at least be treated hypothetically as having occurred). Its features, taken on their own terms, will factor crucially into any initial assessment of its probability. E is, in the language of philosophers, the *explanandum*—the thing to be explained. It requires a good *explanans*—the thing doing the explaining. Within the Law of Conservation of Information, that *explanans* is the event F, the outcome of a search for a search within a search-for-a-search, or S4S, space.

The S4S space is typically much more complicated than the original search space, and so calculating probabilities for it (as with F) will be more complicated than calculating probabilities for the original search space (as with E). Such considerations do not strictly prove that calculating a small probability for E always trumps calculating a large probability of F when the two are in tension. But they do suggest that if there is a good argument for the probability of E being small, then there needs to be a better argument for the probability of F being large if F having a large probability would in turn force E to have a large probability.

In applying the Law of Conservation of Information, we typically have a good handle on the baseline search and the probability it assigns to the event E. By contrast, for the event F associated with the search for a search, we typically infer its probability through a less sure process of reasoning. In practice, then, the Law of Conservation of Information more readily assumes the role of a probability diminisher (leveraging a small probability of E to diminish the probability of F) than a probability amplifier (leveraging a large probability of F to amplify the probability of E).

10.5 Intelligent Agents as Probability Amplifiers

The Law of Conservation of Information, as formulated in this paper, is concerned with identifying search processes that amplify the probability of otherwise improbable events. In this law, there is a standard relation among probabilities, which takes the form of a conditional: If $P(\text{Event})$ is near zero and $P(\text{Event}|\text{Amplifier})$ is near one, then $P(\text{Amplifier})$ is near zero. The mathematics behind this conditional is the by now well-established probabilistic inequality, expressed here as $P(\text{Event}) \geq P(\text{Event}|\text{Amplifier}) \times P(\text{Amplifier})$.

This probabilistic relation is illustrated in the next section with respect to evolvability (Section 11). There the event in question is the origination of a given protein and the amplifier is a particular form that natural selection can take (a particular fitness landscape operating within a particular environment). And so the Law of Conservation of Information requires comparing $P(\text{Protein})$, $P(\text{Protein}|\text{Natural Selection})$ and $P(\text{Natural Selection})$. Provided $P(\text{Protein})$ is small, the Law of Conservation of Information then shows that natural selection, in the particular form it takes to produce the protein, is highly improbable and therefore fails to account for the information in the given protein.

The question may now reasonably be asked: what if in place of natural selection we make the probability amplifier to be an intelligent agent? To put this question in its starkest form, what if the probability amplifier is God? The role of God as a probability amplifier in explaining biological complexity (such as the complexity of proteins) arises because natural selection is commonly viewed among materialists as a God substitute. Richard Dawkins, for instance, remarks, "I could not imagine being an atheist at any time before 1859, when Darwin's *Origin of Species* was published... [A]lthough atheism might have been logically tenable before Darwin, Darwin made it possible to be an intellectually fulfilled atheist" [21, pp. 5–6]. For Dawkins, without Darwin, William Paley's design argument would still win the day,[33] we would be compelled to be theists, and the God hypothesis would have legitimacy throughout the biological sciences. For Dawkins, because of Darwin, natural selection has replaced God.

If we now run the logic of the Law of Conservation of Information for the previous protein example, substituting God for natural selection, we face a disconcerting possibility. Namely, if we find that $P(\text{Protein})$ is near zero and $P(\text{Protein}|\text{God})$ is near one, would this not put $P(\text{God})$, the probability of God, near zero? In that case, how, if at all, could the Law of Conservation of Information distinguish between natural selection and God? And how could it support a case for God, or more generally a case for intelligent design? In fact, it would seem that by finding $P(\text{Protein}|\text{God})$ to be large and by thus diminishing $P(\text{God})$, we would in effect be offering an anti-theistic argument.

A committed theist might now counter that on most theological accounts, God is a necessary being. Thus the probability of God, $P(\text{God})$, would have to be 1. But if $P(\text{Protein}|\text{God})$ is near 1, and if $P(\text{God})$ is 1, then $P(\text{Protein})$ itself would have to be near 1 (because probability theory requires that $P(\text{Protein})$ be greater than or equal to $P(\text{Protein}|\text{God})$ times $P(\text{God})$). In plain English, if God exists necessarily and if God intends to bring about a certain protein, then the probability of that protein existing is near 1, if not exactly 1.[34] But as-

[33] The very title of Dawkins book *The Blind Watchmaker* alludes to William Paley's design argument, which in its most memorable form argues that an intelligent—not blind—watchmaker can reasonably be inferred from finding a watch in a field. See [35, ch. 1].

[34] Depending on one's view of divine sovereignty, the probability of the protein existing can be not just near 1 but exactly 1, as when

signing probabilities at or near 1 to extant proteins seems scientifically indefensible. The problem, as evident in Section 11 (on evolvability), is that functional proteins are sparsely distributed in the space of all possible proteins (polypeptides). Consequently, unless external information is applied—in keeping with the Law of Conservation of Information—proteins taken by themselves need to be assigned low probability, especially given the radical contingency among ways amino acids can be joined, a point made in Subsection 10.4.

I have written at length elsewhere about handling agent explanations within a Bayesian probabilistic framework.[35] I do not want to rehearse that work here except to say that agents, be they God or intelligences generally, do not lend themselves to probabilistic analysis of their creative innovations. What is the probability of me writing this paper? What is the probability of Bach composing his Brandenburg Concertos? What is the probability of Nikola Tesla inventing the alternating current induction motor? Such probabilities put agency (whether human, alien, divine, or otherwise) in the same boat as natural laws, mechanisms, algorithms, mathematical procedures—in a word, search processes. But we have no way to make sense of creative innovation in terms of search processes and thus no way to assign probabilities to creative innovation. If we did, we could predict creative innovation. But we cannot.

This breakdown in probabilities with creative innovation is especially true of God. Atheists and theists could presumably agree that God is often not particularly forthcoming about his intentions. Given a particular protein, who is to say what value $P(\text{Protein}|\text{God})$ should take? Probabilities need to be assigned on the basis of known information. For agent innovation, and especially divine innovation, however, we typically do not know enough to meaningfully assign probabilities. And even if we could meaningfully assign a numerical estimate to $P(\text{Protein}|\text{God})$, perhaps only hypothetically, that still would not help us with $P(\text{God})$.

In a typical Bayesian context, $P(\text{God})$ is the prior probability of the existence of God. But in the context of the Law of Conservation of Information, focused as it is on search, $P(\text{God})$ would be the probability that God is inclined to execute a particular search. Consequently, there would be different God hypotheses for different searches God might be inclined to execute, and thus different probabilities assigned to these different God hypotheses. But who seriously thinks they have enough insight into the mind of God to estimate such probabilities?

In all this, it is important to recognize that nothing fundamentally prevents intelligent agents from acting as search processes under the Law of Conservation of Information. Intelligent agents can certainly mimic search processes, executing algorithms, as it were, by hand. But more commonly, intelligent agents can exhibit large scale probabilistic behavior that is well established and can be relevant to understanding success in search. Take, for instance, performance on the SAT test. Performance in this case is entirely due to agents—the test-takers. Scores range between 400 and 1600 in 10-point increments (400, 410, 420, 430, ..., 1580, 1590, 1600). This, we might say, is our search space. The 50th percentile score is around 1050. The top score of 1600 corresponds to roughly .02 percent of test-takers, and so has a probability of roughly 1 in 5,000.[36]

Suppose now we let Event denote getting a 1600 on the SAT and let Amplifier denote those students that ace the SAT by getting a 1600. Students who take the SAT, we might therefore say, constitute the search-for-a-search (S4S) space. Given these assumptions, $P(\text{Event}) = .0002$ and $P(\text{Event}|\text{Amplifier}) = 1$. And what about $P(\text{Amplifier})$? Amplifier then consists of those students that score 1600 among all 2 million or so students that take the SAT annually (roughly 400 students). And so, $P(\text{Amplifier})$, consistent with the Law of Conservation of Information, will be less than or equal to .0002, which is exactly what we should expect. Indeed, seeing a score of 1600 on the SAT is improbable, and it is improbable because finding students capable of attaining that score is likewise improbable. Moreover, these improbabilities are commensurable in exactly the way required by the Law of Conservation of Information.

The bottom line here is that conservation of information applies to intelligent agents so long as the needed probabilities associated with their role in search can be calculated. Where agents operate in well-defined probabilistic ways to perform searches, the math all works out and the Law of Conservation of Information applies. On the other hand, where agents operate as creative innovators, defying predictability and short-circuiting probabilistic analysis, the math breaks down and the Law of Conservation of Information fails to apply because the numbers simply are not there to make it apply.

11. EVOLVABILITY

Evolvability is about evolutionary interconnectedness—the degree to which evolving things are connected

God always acts with power and finality to achieve whatever God wants, as would be the case in strong predestinarian theologies such as those of Saint Augustine and John Calvin.

[35] See, for instance, [36, chs. 32–34].

[36] On its website, the College Board, which administers the SAT, focuses on SAT percentiles that top out at 99. See https://research.collegeboard.org/reports/sat-suite/understanding-scores/sat. But that 99th percentile leaves a lot of variability. For a higher precision analysis of SAT percentiles, that adds more significant digits to the percentiles and from which the numbers in this paper are taken, see https://blog.prepscholar.com/sat-percentiles-high-precision-2016. Both sites were last accessed on January 31, 2025.

by evolutionary paths. Evolvability came up in Section 8 in the context of displacement, but I want next to treat it in the more general context of conservation of information. I am going to need to do a bit of stage-setting before showing how evolvability relates to conservation of information. The more that evolutionary paths connect to an individual thing, the more evolvable is that thing. The more evolutionary paths connect things considered collectively, the more evolvable is the space made up of those evolving things. One can think of evolvability as a directed graph where the nodes are things undergoing evolution and where each edge is an evolutionary transition (think of an arrow) connecting a start node to an end node [37, pp. 5–6]. An evolutionary path or pathway is then a sequence of edges where the end node of a prior edge is the start node of the next edge.

In this picture of evolvability, at one extreme are nodes with no edges at all connecting them, thus indicating complete isolation and no evolutionary connection whatsoever. At the other extreme are nodes with edges in both directions connecting all pairs of nodes, indicating not just complete evolutionary connectedness, but even instantaneous evolution of any one node into any other. In between are levels of evolvability that require longer evolutionary pathways consisting of multiple edges to connect nodes. Evolvability typically exhibits a mixture of connectedness and isolation. Individual nodes or regions of nodes may be connected by pathways to other nodes or regions of nodes whereas other nodes or regions of nodes may be unconnected in this way and therefore isolated.

With evolvability, there is always an interplay between two things: (1) an evolutionary mechanism that defines the permissible evolutionary paths; and (2) a space whose items may evolve according to this evolutionary mechanism. The evolutionary mechanism can be thought of as a set of transition rules that tell us which steps in the evolutionary process can be taken and which are proscribed. The rules can be generous, giving a wide berth to evolvability. They can be miserly, severely limiting evolvability. They can be somewhere in between.

As for the space over which evolution takes place, it need not be limited simply to functional items (such as extant organisms in the case of biology). Typically it takes the form of a *configuration space*. Such a space consists of all relevant possible configurations, only some of which will be evolutionarily significant (significance being cashed out in terms of meaning, function, viability, fitness, etc.). Such a space will therefore include disconnected items that are evolutionary isolates into which and out of which no evolution can occur via the posited evolutionary mechanism. To the degree that we have a handle on such spaces, we can assess not just where evolution is likely to succeed but also where it is likely to fail. Items in the space that cannot reasonably belong to an evolutionary path need as much to be factored in as those that can.[37].

In the context of written language, a configuration space typically takes the form of all possible letter sequences. Evolution within such a space could then be concerned with the evolutionary connections between meaningful letter sequences. In the context of biological evolution, a configuration space typically takes the form of different ways of assembling basic chemical constituents. For instance, a configuration space of DNA sequences consists of all possible nucleotide sequences. Any evolution within such a space is then concerned with evolutionary connections between sequences that are biologically meaningful, such as DNA sequences that code for proteins or assist in gene regulation.

The question of evolvability is central to Darwinian evolution. Darwinists take a gradualistic view of biological evolution in which all organisms can be connected via small evolutionary steps—let us call them baby steps—each of which is selectively advantageous or at least neutral. Evolutionary pathways connecting distinct organisms via baby steps may be direct or issue from an earlier common ancestor. Because Darwinism understands all organisms as connected to each other in this way, universal common ancestry is baked into Darwinian theory. But such universal connectedness via gradualistic evolutionary pathways cannot simply be presupposed—it must be verified. Moreover, even if universal common ancestry holds, it is a separate question whether it happened by a specifically Darwinian form of evolution that gives pride of place to natural selection. In any case, the evolvability required for Darwinian evolution to work requires evidence. Conservation of information, as we will see, is a tool for evaluating such

[37] Cognizance of evolutionary dead ends has a long history. Take, for instance, the following passage from the Roman poet Lucretius (ca. 99 to 55 BC):

> Many were the monsters also that the earth then tried to make ... [that] could do nothing and go nowhere, could neither avoid mischief nor take what they might need. So with the rest of like monsters and portents that she made, it was all in vain: since nature denied them growth, and they could not attain the desired flower of age nor find food nor join by the ways of Venus. For we see that living beings need many things in conjunction, so that they may be able by procreation to forge out the chain of the generations... And many species of animals must have perished at that time, unable by procreation to forge out the chain of posterity: for whatever you see feeding on the breath of life, either cunning or courage or at least quickness must have guarded and kept that kind from its earliest existence...

Lucretius is here clearly addressing evolutionary dead ends to which and out of which no evolutionary paths lead. Quoted from [38, pp. 837–877].

evidence.

Most advocates of biological evolution (Darwinian or otherwise) admit no real limits on evolvability. Sure, there is J.B.S. Haldane's widely reported pseudo-concession that evolution would be falsified if a fossilized rabbit were found in the Precambrian. His point was that evolution, because it is gradualistic, would not allow a rabbit to be directly evolvable from Precambrian organisms.[38] But given enough time for evolution to act, and without restricting evolution to any particular geological period, a suitable evolutionary path always exists to account for rabbits or any other extinct or extant organisms. Haldane was therefore far from putting real-life evolvability on the evidential chopping block. Whatever factors may constrain evolvability, most evolutionists do not admit any insuperable obstacles to the evolutionary interconnectedness of any actual organisms. Thus, for any two organisms, there is always some evolutionary path that connects them, whether directly or through some common ancestor.

In short, evolutionists are convinced that any constraints on evolvability are never so stringent as to prevent evolution from finding a way around them. Accordingly, evolutionists treat evolvability as always working in their favor, presenting no impassable barriers and underwriting all the evolutionary change that their theory of evolution requires. But evolvability need not be a friend of evolution. Indeed, evolving systems can face stark limits on the evolutionary change that is possible. For one thing to evolve into another thing that is quite different from it, the one might simply change into the other immediately by a macroevolutionary event—essentially by a miracle. Since evolutionary theory frowns on miracles, the other possibility is preferred—namely, that the one evolves into the other by a gradual evolutionary process consisting of a sequence of functional intermediates each of which is very similar to its immediate predecessors and successors. Stretch out the evolutionary process into such a sequence of baby steps, and what can evolution not accomplish? But conversely, what can evolution accomplish if the needed functional intermediates simply do not exist? The short answer to this last question is, Nothing.

To see how evolution can break down apart from the right and needed functional intermediates, let us for the moment set aside biological evolution and consider the evolution of English words. English words could be said to "evolve" into other English words by the transformation rules (evolutionary mechanism) of:

1. adding a single letter,
2. deleting a single letter, or
3. substituting a single letter,

where each of these letter changes must yield a meaningful word. All changes are thus gradual in that they allow only a single letter to change at any step in the evolutionary process. Allowing multiple letter changes at a step would lead to a less gradual form of evolution and to greater evolvability (such as allowing any combination of two or fewer letters at time to be added, deleted, or substituted all the while preserving meaningful words).

So here is the question: how much evolution is possible in this configuration space of English words when subjected to these three transformation rules, which allow only single-letter changes at a time and require all changes to result in meaningful words? Obviously, the answer will depend on the size of the English dictionary over which this form of evolution takes place—the larger the dictionary, the more evolution is possible, and thus the greater the evolvability. To facilitate English-word evolution, let us therefore be generous in our choice of dictionary by using Unix Words, which has over 200,000 English words. Most spellcheck dictionaries are considerably shorter, coming in around 90,000 words.[39]

Given the transformation rules of adding, deleting, or substituting a single letter at a time while always retaining meaningful words, the word A (the indefinite article) can evolve into the word TIPS. Thus one evolutionary path from A to TIPS is A → AN → TAN → TIN → TIP → TIPS. But this evolutionary path from A to TIPS is not unique. Another path with the same starting point and endpoint, and using the same transformation rules, is A → AT → BAT → BATS → BASS → BOSS → LOSS → GLOSS → GROSS → DROSS → DROPS → DRIPS → DIPS → TIPS.

On the other hand, using only Unix Words, an exhaustive computer search shows that running through all possible evolutionary paths starting from the word A, evolution cannot reach the words HYMN, ENVY, TOFU, OBEY, and IDOL. These are just a few of the four-letter words into which the word A cannot evolve. Longer words such as CONSTANTINOPLE are also provably unreachable via this form of evolution from the word A.

[38] "Haldane once said he would give up his belief in evolution if someone found a fossil rabbit in the Precambrian. The reason is that the rabbit, which is a fully formed mammal, must have evolved through reptilian, amphibian, and piscine stages and should not therefore appear in the fossil record 100 million years or so before its fossil ancestors." Quoted from [39, p. 66].

[39] For Unix Words, see https://en.wikipedia.org/wiki/Words_(Unix). Optimal spell checkers weigh in around 90,000 words, thus well below Unix Words' 200,000. Going over 90,000 invites too many false negatives, which is to say words that are actually misspelled but that are treated as spelled correctly because they are at once rare but also similar to commonly misspelled worlds. For instance, "baht" denotes the basic monetary unit of Thailand. But among English speakers it will be so rare that its appearance in most documents would probably mean that the word "bath" was misspelled. See https://en.wikipedia.org/wiki/Spell_checker. Both these entries were last accessed on February 22, 2025.

This example therefore illustrates that forms of evolution exist that are demonstrably limited, with evolutionary paths between functional members of the configuration space being provably nonexistent.

Consequently, with English-word evolution, we can know for a fact that the structure of the evolutionary configuration space together with the evolutionary transformation rules acting on it, when taken jointly, limit evolvability between words. Such a limitation on evolvability can lead to isolated islands of function (meaning) that cannot be evolved into or out of (such as no evolutionary path existing from A to IDOL). In place of isolated islands of function in the evolutionary configuration space, greater evolvability (such as by allowing multiple letter changes at once) would lead to greater evolutionary connectedness. Sufficiently lenient transformation rules could even allow complete evolutionary connectedness, with pathways of functional intermediates (meaningful words) connecting all otherwise isolated islands.

Now it might be argued that the constraints on evolvability of Unix Words as just described will tend to underestimate evolvability. The argument would be that in the formation of words, humans can, as a matter of convention, coin any words they like by arbitrarily investing with meaning any letter combinations whatsoever. Unix Words is a big dictionary as dictionaries go. But to keep up to date, it needs to be dynamically changing. So what count as words today will change (evolve) in the future as human use of the English language changes. As newly coined English words are added to Unix Words, evolvability over the space of letter sequences will increase.[40]

Interestingly, the argument just given to expand evolvability over letter sequences because humans can always invent and add new words or acronyms does not carry over to biology. Arbitrary letter combinations, because they can be invested with meaning as a matter of human convention, can always become meaningful (especially if we allow acronyms). But arbitrary combinations of chemical constituents, because of their intrinsic chemical properties, may rule out all biological function (viability). Peptide sequences of amino acids, for instance, need to adopt specific conformations and dynamic states to carry out their biological functions. This is true whether they are fully structured folded proteins or intrinsically disordered proteins. Protein function reflects chemical principles that constrain proteins depending on composition and interactions of their chemical constituents. Thus, unlike human convention, which can invest any letter combinations with meaning, chemistry imposes strict limits on the chemical combinations that can exhibit biological function.

Limitations on evolvability over biological configuration spaces might therefore exceed limitations on evolvability over linguistic configuration spaces. Whereas linguistics allows that any combination of letters could be meaningful or have a function, biology, by being grounded in chemistry and physics, can definitively ensure that certain chemical combinations in the biological configuration space lack function. This is not to say that biological configuration spaces must as a matter of definition suffer from limited evolvability among their biologically functional elements. But it is to say that, as in the English-word evolution example, they might suffer from limited evolvability and with even more reason.

Of course, the point of interest with biological configuration spaces is to get a convincing handle on evolvability within them. Let us therefore turn to the configuration space of polypeptides and try to understand evolution in this space of proteins and protein folds. Consider, therefore, Douglas Axe's research on beta-lactamase, an enzyme granting bacteria antibiotic resistance. Rather than focusing explicitly on the evolvability of beta-lactamase from some other enzyme, Axe set out to determine how rare functional beta lactamase sequences are within a tiny region of sequence space marked by a cluster of closely related natural beta-lactamases. He found that even in this highly promising patch of sequence space, functional sequences are extraordinarily rare—something like one working sequence among a staggering 10^{77} non-functional ones [40].

In other words, for an undirected process starting with a protein that is outside the tiny beta-lactamase region of sequence space to *happen upon* a working beta-lactamase is equivalent to solving a very difficult search problem—a molecular needle-in-a-haystack problem. But, as established here, the Law of Conservation of Information shows this to be effectively impossible. Thus, however much Darwinian critics of Axe's result would like to think that evolvability and rarity are separate problems, they are actually the same problem viewed from two angles: extraordinary rarity of the kind demonstrated by Axe necessarily implies a lack of evolvability.

Darwinian critics have argued that Axe's calculations are flawed because they fail to account for unknown evolutionary pathways that might have led to beta-lactamase. In particular, they argue that functional shifts through co-option—where proteins and folds initially serving one role are later repurposed for another—could explain beta-lactamase's emergence. Yet, such indirect evolutionary pathways require a highly specific sequence of events where both structural transformations and functional shifts have to align. Such an alignment is required by Darwinian evolution, but to be credible it also requires independent evidence, and not just that Darwinism needs it. And invariably, for biological systems of any com-

[40] For the complete list of Unix Words, see https://raw.githubusercontent.com/dolph/dictionary/refs/heads/master/unix-words (last accessed February 24, 2025).

plexity, such as beta-lactamase, that evidence is lacking. The appeal to co-evolution and co-option in such cases always ends up being a just-so story.

In any case, Darwinian critics of Axe insist, as their theory demands of them, that functional proteins must have evolved through small incremental steps, preserving function at each stage, even if the exact pathways remain unidentified. This assumption, however, shifts the burden of proof away from Darwinists, allowing them to reject small probability calculations outright while accepting large probability calculations when they support natural selection. This double standard invalidates their critique, as they demand from design theorists a degree of rigor in probability calculations that is impossible to attain while at the same time they exempt themselves from providing alternative calculations that show how beta-lactamase could reasonably have evolved by Darwinian means.

Ultimately, the question of beta-lactamase evolvability depends on structural, and specifically biophysical, features of the polypeptides that are evolving as well as on the evolutionary mechanism that is constraining their evolution. Evolvability in this instance therefore depends on

1. the structural constraints on protein folding;

2. the density distribution of functional sequences within the biological configuration space of all possible polypeptide sequences (are the sequences sparse or common?—all the evidence suggests that they are sparse [41]); and

3. the degree to which the underlying evolutionary mechanism is able to connect functional polypeptide sequences by gradual evolutionary paths each of whose intermediates is functional.

The evolvability of beta-lactamase from potential evolutionary precursors via Darwinian processes has an objective probability (or at least range of probabilities) that, given enough experimental power (power that we possess with English-word evolution though not with biological evolution), could be precisely determined. Axe's work represents a rigorous attempt to estimate this probability. Dismissing Axe on the basis of a Darwinian faith in evolutionary pathways that are unknown, perhaps unknowable, and possibly nonexistent, does not constitute a scientific rebuttal of Axe's work. Proper evaluation of this work requires engaging with the biophysical realities of protein evolution and the probabilistic hurdles they face rather than simply assuming that viable Darwinian pathways must invariably exist.

Axe's work presents an argument from small probabilities. Given the functional specification of beta-lactamase in facilitating antibiotic resistance, it is straightforward to reframe his argument as demonstrating the specified complexity of beta-lactamase and thus, because specified complexity is a reliable marker of design (see Section 7), to conclude that beta-lactamase is designed. I am happy with this conclusion and regard the case for the specified complexity of beta-lactamase as strong. As it is, Winston Ewert and I reframed Axe's argument as a specified complexity argument in the second edition of *The Design Inference* [19, pp. 372–378].

But suppose you are a Darwinist who is convinced that the Darwinian mechanism has a way to bring about beta-lactamase apart from design. You may have no rebuttal to Axe's small probability argument, but you remain unconvinced. In Bayesian terms, perhaps your prior probability of Darwinism being true is so great that it overwhelms any design hypothesis, however well supported it might be. Perhaps you think design explanations so totally violate parsimony (Occam's razor) that they cannot be permitted on first principles. For instance, biologist Bret Weinstein, in a podcast with Joe Rogan, will admit, "I think modern Darwinism is broken," adding that its mechanism "isn't powerful enough to explain the phenomena that we swear it explains."[41] And yet he sees parsimony as necessitating something like Darwinism and therefore as blocking design:

> If you take the intelligent design folks and you extrapolate from what they seem to be suggesting, they do not escape a necessity for a Darwinian explanation. Even if the creatures of Earth were designed on a drawing board by a creature that wanted to make them, that creature has to have come from somewhere. And the only explanation that has ever been proposed for where such a creature could have come from is Darwinian evolution. So to me, the problem with intelligent design, the most fundamental one, is that even if it were true, you have solved the problem of explaining Earth's creatures at a cost that is a million times worse in terms of parsimony. If it is hard to explain a tiger through Darwinian processes, it is that much harder yet to explain a tiger designer.[42]

This is the old who-designed-the-designer objection. It has force if you are a materialist because in that case any evolved creature must in turn be explained by some other creature. And eventually explaining a creature by evolution from another creature must end in some creature that requires no explanation, such as a primordial goo at the origin of life. So the committed materialist

[41] Joe Rogan interview with Bret Weinstein, number 2269, February 6, 2025, available at https://www.youtube.com/watch?v=7ted-qUqqU4 (last accessed February 27, 2025).

[42] Ibid.

needs something like Darwinism to explain the development of life once it is here, and also needs a materialistic story for how life could originate in the first place. Design, of course, allows that what is causally behind any creature, evolved or otherwise, may not be a purely material creature but could instead be an ultimately nonmaterial information source for which the creature is an informational receiver. I address Weinstein's objection and why it fails at length in the next section, putting the objection there in the mouth of Richard Dawkins.

In offering his reservations about both Darwinism and design, Weinstein is tacitly invoking conservation of information. Explaining a tiger by explaining a tiger designer is a regress characteristic of conservation of information, which guarantees that the information problem only gets worse as the regress is taken further back. But note, from the vantage of intelligent design, a tiger designer does not require the same sort of explanation as a tiger. The tiger designer could be a different sort of entity from the tiger—namely, a nonmaterial irreducible intelligence—a possibility considered in the next section on the information regress problem (Section 12) and explored further in the conclusion of this paper (Section 14).

Returning to Douglas Axe's work on beta-lactamase, design theorists might therefore try to hold Weinstein's feet to the fire by arguing that the probability of its occurrence by Darwinian means is indeed very small. Accordingly, Axe has discovered a clear case of specified complexity, and thus, by the logic of the design inference, is warranted in inferring that beta-lactamase is designed. But the design theorist has another option, which is to concede for the sake of argument Weinstein's claim of universal Darwinism—namely, that ultimately some Darwinian-like process had to be responsible for all biological creatures. Thus, in particular, Axe and fellow design theorist could, for the sake of argument, concede that the Darwinian mechanism might raise the probability of beta-lactamase sufficiently (how, no one knows) to rule out a design inference based on specified complexity. And then, having made this concession, they would still show that Darwinism cannot dispense with design on account of conservation of information.

This, finally, is where conservation of information fits in relation to evolvability. For the Darwinian mechanism to explain beta-lactamase, it must raise the probability of its being formed. Moreover, by considerations of radical contingency and multiple degrees of freedom, as described in Subsection 10.4, the baseline probability of this system is extremely low inasmuch as biologically functional polypeptides are extremely rare within the totality of polypeptides. So somehow the Darwinian mechanism, or some other material mechanism capable of propelling biological evolution, must have raised the baseline probability to that of an improved probability.

But now the Law of Conservation of Information applies, showing that the very mechanism that could give rise to beta-lactamase (or any other highly improbable target—there is nothing special in the argument here about beta-lactamase) must itself be improbable. The improbability of beta-lactamase, even under Darwinian or other material mechanisms, is thus shifted to the improbability of the mechanisms and the conditions under which they are acting. The move here exemplifies the classic shift in the Law of Conservation of Information from an initial improbability of a target in a search (the baseline probability) to a later improbability of a higher-level target in a search for a search (the S4S probability). As in all applications of the Law of Conservation of Information, the very improbability that material evolutionary mechanisms were supposed to resolve now returns with greater force than before.

The Law of Conservation of Information thus applies to any evolutionary scenarios in which the functional things that could be the product of evolution are sparsely distributed and therefore highly improbable against the backdrop of the underlying biological configuration space. The law implies, as should now be clear, that even when evolvability is high and supported by specific mechanisms, this does not overturn design or suggest those mechanisms are undesigned. In fact, conservation of information does not require a precise characterization of the evolutionary mechanisms that may be active. It simply requires a reasonable grasp of the underlying probabilities. As a probabilistic inequality involving a baseline, an improved, and a search-for-a-search probability, the law holds with perfect generality.

This ineliminability of design from evolvability was clear even from earlier conservation of information theorems that my colleagues and I proved around 2010 (see Appendix 2). Yet Darwinian critics, whether for not understanding conservation of information, for willfully misunderstanding it, or for simply desiring to derail its force, have for more than two decades now engaged in a *fallacy of false equivalence*, which is the fallacy of pretending that two distinct things are essentially the same when they are not. Thus they mistakenly equate no free lunch with conservation of information. So when they resist the application of conservation of information to Darwinian evolution, in fact they are merely resisting the application of no free lunch to Darwinian evolution.

By themselves, the no-free-lunch theorems do not refute Darwinism, nor have I ever claimed that they do. Yes, I wrote a book titled *No Free Lunch* (published in 2002) that used the no-free-lunch theorems as a springboard to critically re-examine the power of the Darwinian mechanism in explaining biological complexity and from there to argue that design is a better explanation of biological complexity. Yet Darwinian critics of conservation of information have ever since conflated conservation of information with no free lunch, arguing (rightly) that no free lunch fails to refute Darwinism, but then claiming (wrongly) that conservation of information is identical with no free lunch and thus likewise fails. Hence the fallacy of false equivalence, in which these critics refute what was not contested and then claim that the refutation refutes something else.

Joseph Felsenstein is a case in point, but he is representative of Jason Rosenhouse, the late Mark Perakh, Richard Wein, Erik Tellgren, Ole Häggström, and the duo of Wesley Elsberry and Jeffrey Shallit.[43] The no-free-lunch theorems show that any algorithm averaged over fitness landscapes (Felsenstein calls them "fitness surfaces") is equivalent to any other algorithm when averaged in this way. It then follows that any such averaged algorithm is equivalent to blind search. In the statement of these theorems, no restriction is placed on the fitness landscapes. Thus they can be completely jagged, with one member of the configuration space having a completely different fitness value from an immediate neighbor. Most fitness landscapes would thus be like hash functions in cryptography (Felsenstein describes them as "white noise fitness surfaces") where the slightest change in an input completely alters the numerical output [42].

Now Felsenstein's point is that such jagged fitness landscapes, as allowed by no-free-lunch theorems, violate biological reality. The fitness landscapes conducive to evolvability, such as might characterize how genotypes promote reproductive success, must, according to him, be smooth. Why? Because of physics. Thus he remarks:

> [I]t seems reasonable that physics itself predisposes fitness surfaces to be smooth. After all, the forces of physics tend to be local in time and space and to lose influence with greater distance. If I gesture with my fingers, there is hardly any effect on objects at a distance. The floor of our room does not collapse, nor does the ceiling cave in. Different objects are scarcely affected by each other.[44]

The problems with this reasoning are numerous. In the first place, physics does have a place for precipitous and even nonlocal effects. Quantum entanglement is instantaneous and unaffected by distance. Phase transitions in nonlinear dynamics may produce radical changes in the behavior of a system with but the slightest change in a parameter. Felsenstein's argument is based on folk physics, not modern physics and as such is unconvincing.

Leaving aside physics, Felsenstein cites mutational effects as empirical evidence for smoothness of fitness landscapes. In fact, he defines smoothness to mean that small mutational changes lead to small functional changes. According to Felsenstein, the empirical evidence shows that single mutations typically result in small, predictable changes rather than chaotic, random effects. But smoothness in this sense admits striking exceptions. A host of serious genetic diseases such as sickle cell anemia, phenylketonuria, and Tay-Sachs disease all result from single point mutations. Nonsense mutations, where a single nucleotide substitution produces a stop codon within a coding sequence, cause premature termination of protein synthesis, sometimes with lethal consequences. Frameshift mutations, which may be caused by a single nucleotide insertion or deletion, completely disrupt protein synthesis downstream of the mutation, often with huge consequences. And even single amino-acid substitutions caused by single nucleotide substitutions can have drastic effects, depending on the location and the role of the protein. Thus, contrary to Felsenstein's smoothness hypothesis, minor genetic change can and does lead to a catastrophic loss of function.

According to Bill Gates, "DNA is like a computer program but far, far more advanced than any software ever created" [43, p. 188]. But, as we just saw, smoothness admits exceptions with digital codes in cells. Likewise, it admits exceptions with digital codes in computer software. In fact, it is fairer to say that smoothness is the exception rather than the rule in computer programs. Indeed, one stray letter, one comma out of place, can subvert a program, preventing it from achieving its objective, if it even functions at all.

Evolutionists never define smoothness of fitness landscapes with anything like mathematical precision. But let us say that smoothness could be made mathematically precise and that for biological evolution to work it requires smooth fitness landscapes. Smoothness, even then, cannot guarantee the evolvability needed for Darwinian theory to be successful. Smoothness has to mean something like that by moving to near neighbors, fitness should not change drastically. Empirical evidence from some genetic diseases, as noted, contradicts smoothness in this sense. But even if smoothness could be said to hold most of the time, it says nothing about regions of a biological configuration space where fitness is zero. Indeed, fitness landscapes that are zero over such regions are as smooth as they can be over those regions. Moreover, if fitness is zero over sufficiently vast regions of a configuration space, the viable candidates for evolution will be sparsely distributed, drastically curtailing evolvability.

Additionally, smoothness does nothing to assuage the problem of multiple local optima where fitness cannot be improved by gradual evolution. In this case, evolution

[43] See Joseph Felsenstein, "Conservation of Arguments," *Panda's Thumb* (February 6, 2025): https://pandasthumb.org/archives/2025/02/Conservation-of-arguments.html (last accessed February 24, 2025). For the criticisms by Rosenhouse et al., see the relevant links in the Felsenstein article.

[44] Felsenstein, "Conservation of Arguments."

gets stuck because it cannot break away from those local optima, there being no gradual way for fitness to improve. Fortunately, evolution allows for more radical genetic events than point mutations, such as inversions, translocations, deletions, and duplications. But in that case, smoothness loses its hold because, with such large-scale structural changes, all bets are off as to what effect they may have on evolution. Smoothness is essentially a continuity condition, and continuity is always a matter of small local changes in an independent variable leading to small changes in a dependent variable. Smoothness says nothing about how the dependent variable may change in response to large-scale non-local changes in the independent variable. And genetics allows such changes.

Felsenstein's argument can thus be summarized as follows:

1. Real-life biological evolution requires smooth fitness landscapes.
2. Physics and chemistry guarantee that real-life fitness landscapes are smooth.
3. Any probability estimates of smooth fitness landscapes among the totality of fitness landscapes are therefore irrelevant.
4. No-free-lunch theorems hold for the totality of fitness landscapes, both smooth and non-smooth.
5. No-free-lunch theorems therefore do not apply to real-life biological evolution.
6. Conservation of information is equivalent to no free lunch, so conservation of information is likewise irrelevant to real-life biological evolution.

Felsenstein's argument is problematic for more than just point 6. As we've seen, smoothness as applied to fitness landscapes in evolution is questionable on both theoretical and empirical grounds. Points 1 to 3 are therefore problematic. But even granting points 1 to 5, the falsity point 6 means conservation of information puts design back in the running in biology. Even though Felsenstein mistakenly conflates of the two, conservation of information is distinct from no free lunch. Whatever fitness landscapes and other conditions on evolvability are needed to bring about a system like Douglas Axe's beta-lactamase, if those conditions raise the probability of success of evolving such a system, those conditions themselves become highly improbable and require explanation.

The baseline probability for a system like beta-lactamase is indeed small because of the radical contingency and numerous degrees of freedom of its peptide constituents (see Subsection 10.4). Evolvability therefore grinds to a halt unless it can overcome these otherwise inherent improbabilities (just as, in Section 8, Dawkins' hill-climbing algorithm overcame the inherent improbability of generating by purely random means the target phrase METHINKS IT IS LIKE A WEASEL). Conservation of information tracks the probabilistic cost of overcoming improbability. As such, conservation of information is not concerned with the details of how evolution does it. In particular, it is irrelevant if smooth fitness landscapes, however these are defined, are improbable within a wider class of fitness landscapes, and whether a no-free-lunch theorem exists or does not exist to tease apart the fitness landscapes that facilitate evolution from those that do not.

The fundamental problem with Felsenstein's critique of conservation of information is this: it is simply irrelevant whether no-free-lunch theorems undercut or fail to undercut Darwinian evolution. What is relevant is that no-free-lunch theorems have no way to undercut conservation of information. Conflating the two is illegitimate. True, in the history of scientific discovery, no free lunch preceded and provided inspiration for conservation of information. No free lunch showed that when averaged over fitness landscapes, no search algorithm outperformed any other. But some search algorithms do outperform others for given problems. What, then, empowers some algorithms to outperform others? The short answer: information.

No free lunch raises the right question but by itself does not answer it. The answer depends on conservation of information. Conservation of information characterizes the performance of algorithms in terms of their probability of successfully carrying out searches (the bigger the probability, the better the performance). The very statement of the Law of Conservation of Information, by being framed strictly in terms of probabilities, contains no reference to fitness landscapes and in no way depends on the statements or proofs of the no-free-lunch theorems. That is not to say that fitness may not arise in some applications of conservation of information. Appendix 2, for instance, in recapping previously proven conservation-of-information theorems, states a fitness-theoretic conservation-of-information theorem. But in any discussions of evolvability, conservation of information can be considered on its own terms irrespective of no free lunch.

12. THE INFORMATION REGRESS PROBLEM

The Law of Conservation of Information raises the prospect of a regress that was touched on in Section 4. There, in the treasure hunt example, the treasure had a very small probability, call it p (the baseline probability), of being successfully found on its own. Consequently, the treasure hunter went to the offices of the mapmaker Rand McNally to find a map that with substantially bigger probability, call it q (the improved probability), would help find the treasure. But finding such a map, according to the Law of Conservation of Information, has probability no more than p/q, which sets an upper bound on the probability of a successful search for a search (the S4S probability).

Undeterred by the failure of this search for a search to raise the probability of successfully carrying out the original search, the treasure hunter might now try to find a way to locate a successful search for a search for a search (S4S4S), in effect bootstrapping the problem of finding the treasure to higher levels of search. Thus, the treasure hunter might visit the offices of a search index company (a subdivision of Google, perhaps) that can provide instructions to find a map at the offices of Rand McNally that in turn can successfully find the treasure. The search index company might thus give instructions such as, "At the Rand McNally offices, take the elevator to the fifth floor, take a right and enter the fifth door on the left, and then look at the hundred and seventh filing cabinet, open the bottom drawer, and take out the eighty-ninth map counting from the front." But the search index company will have a plethora of such instructions, most of which will not yield a successful search for a search, and so will not help with the original search. Bottom line: just as conservation of information applies to the S4S, it applies to the S4S4S, the S4S4S4S, and so on indefinitely.

All of this is intuitively obvious. Yet a straightforward application of the fundamental inequality that characterizes the Law of Conservation of Information, $P(E) \geq P(E|F) \times P(F)$, also makes this clear mathematically. I will focus on the S4S4S (the search for a search for a search), but the same reasoning applies to indefinite orders of higher-level search (S4S4S4S, S4S4S4S4S, etc.). Suppose E is highly improbable when this event is left to itself ($P(E) = p$), and that F renders E highly probable ($P(E|F) = q \gg p$). $P(F)$, by our fundamental inequality, is now less than or equal to p/q. But suppose next that G (some still higher-level search) renders F highly probable, say with probability r ($\gg p/q$). It will now follow that $P(E) \geq P(E|F) \times P(F)$ and $P(F) \geq P(F|G) \times P(G)$, from which it follows that $p/(q \times r) \geq P(G)$.

Now, if r is strictly less than 1, then $p/(q \times r)$ will be strictly greater than p/q, and so, even though $P(G)$ will still be small, its upper bound of $p/(q \times r)$ will be greater than the upper bound of p/q for $P(F)$. But the force of the regress inherent in the Law of Conservation of Information consists not in $P(G)$ needing to be less than or equal to $P(F)$, or in $P(F)$ needing to be less than or equal to $P(E)$. The Law of Conservation of Information does not require these individual inequalities to hold. Instead, the law implies that from unwinding these probabilities in relation to the conditional probabilities $P(E|F)$ and $P(F|G)$, a search for a search for a search (S4S4S) makes it even less likely to successfully complete the original search than a search for a search (S4S). And still higher levels of search only further exacerbate the problem of successfully completing the original search, never offering any net probabilistic benefit. In consequence, trying to regress out of the Law of Conservation of Information merely intensifies its grip.

The Law of Conservation of Information offers no search benefit via such a regress because the information needed to explain successful search always becomes more intensive the higher up we go in the search hierarchy. Think of the search hierarchy as follows. There is the original search space Ω with a target T that needs to be found. There is the search for a search (S4S) space $\overline{\Omega}$ consisting of all possible searches of the original search space Ω, where the task is now to find a higher-level target $\overline{T}$ within $\overline{\Omega}$. There is the search for a search for a search (S4S4S) space $\overline{\overline{\Omega}}$ consisting of all possible searches of searches ($\overline{\Omega}$) of the original search space Ω, where the task is to now find a still higher-level target $\overline{\overline{T}}$ within $\overline{\overline{\Omega}}$. And so on. Regressing up this search hierarchy would be productive if successfully concluding the original search for T could be made increasingly probable by continually moving up the search hierarchy. In that case, at some point the search for T might be rendered sufficiently probable so that T would stand a reasonable prospect of being found. The Law of Conservation of Information eliminates that possibility, making T increasingly less probable of being found by moving up the search hierarchy.

Not all regresses are bad in the sense that they continue on indefinitely or make problems increasingly difficult to solve. In mathematics, recursion is a productive type of regress where at each stage a simplification occurs and gets us closer to the solution of a problem. Take what is called "n factorial," which is written as $n!$—n followed by an exclamation sign. The definition of this number is commonly written as $n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$ (i.e., multiply together all the integers between n and 1). But this number can also be defined recursively as $n! = n \times (n-1)!$ with the proviso that $1! = 1$. Thus n factorial is written as the number n times the factorial of $n - 1$. By the same token, $n - 1$ factorial is then written as the product of $n-1$ times $n-2$ factorial. And so on. Factorials of smaller numbers are more quickly unwound by this recursive/regressive process than larger numbers. So at each recursive stage, the factorial to be calculated is easier than the one before. Eventually this recursion/regress stops when the process reaches 1! because 1! is by definition just 1. In the process, all the factors between n and 1 get multiplied, yielding a precise calculation of $n!$

If recursion in mathematics constitutes a productive regress, higher-level search subject to the Law of Conservation of Information constitutes an unproductive regress. As we regress up the search hierarchy, the

Law of Conservation of Information in fact guarantees that the original search is being rendered increasingly improbable. It is not possible to get around this law by invoking ever higher levels of search. The Law of Conservation of Information, where it applies, rules out any productive informational regress. This point needs to be underscored because evolutionary theory hides in a so-called productive regress.

Conventional evolutionary theory claims that biological information gradually accrues over a long and unguided evolutionary process, where each step in the process adds a small amount of information that over time cumulates and becomes substantial. Accrue as it might, none of this information can rightly be regarded as created through an evolutionary process but only as re-expressing already existing information—that is, if the Law of Conservation of Information applies to evolutionary theory. Wherever the Law of Conservation of Information applies, outputted information can at best maintain parity with inputted information and often, because of frictional forces, may incur strict losses of information. In terms of search, when the law applies, bringing in higher-level searches cannot improve on the original search. When the law applies, such higher-level searches never yield a productive regress.

The big question, therefore, is the applicability of the Law of Conservation of Information to evolutionary theory and, in particular, to evolutionary biology's information problem. Most evolutionists, in their scientific theorizing, are materialists, seeing the evolutionary process as unfolding through the activity of purely unguided material forces (notably natural selection acting on random variations). Most evolutionists also admit that life has become more complex over the course of evolutionary history, requiring palpable increases of information over time. Nineteenth-century Scottish zoologist Robert Edmond Grant expressed this progressive view of evolution memorably in the phrase "monad to man."[45] Accordingly, life becomes increasingly complex and information-rich as it progresses from simple organisms ("monads") to complex beings ("man"). Most evolutionists, if using the language of complexity and information, would say that evolutionary processes create, rather than merely redistribute, ever increasing orders of biological information. Evolutionary biology's information problem is therefore to explain such increases in information, demonstrating that evolutionary processes do indeed create novel information.

To understand how the Law of Conservation of Information applies to evolutionary biology's information problem, we first need to be clear what sort of information we are talking about when it comes to biology. There is a sense in which information as it unfolds by physical laws is strictly conserved, neither increasing or decreasing. I address this point at greater length in Appendix 1, where I contrast conservation of information in physics versus conservation of information in search. As shown there, physical laws take the form of equations that, via initial and boundary conditions, fully specify future states in terms of prior states and vice versa (the equations of physics being, in general, time-reversible). Future and prior states are therefore recoverable from each other, and so neither state can contain more information than the other. So, if biological information is understood solely from the point of view of physics, biological information could not be said to increase over time, or, if it does increase over time, it does so locally, drawing information from elsewhere in the universe, implying a corresponding decrease in information elsewhere.

Perhaps the most striking number in Roger Penrose's book *The Emperor's New Mind* is $10^{10^{123}}$, which can be interpreted as gauging the universe's total information [45, pp. 343–344]. For Penrose, $10^{10^{123}}$ is the jaw-droppingly vast number of possible initial configurations the universe could have taken at the time of the Big Bang. Calculated from the Bekenstein-Hawking entropy associated with the universe's total mass-energy and its cosmological horizon, this number captures the extraordinarily low-entropy state of the early universe, which is necessary to account for the arrow of time and the emergence of order. Penrose uses this number to argue that the initial state of the universe was extraordinarily fine-tuned so that out of $10^{10^{123}}$ possible states, the actual one occupies an extraordinarily tiny fraction of this immense possibility space. The reciprocal of $10^{10^{123}}$ (i.e., $10^{-10^{123}}$) can therefore be interpreted as the probability of getting a universe as finely tuned ours. Its exponent, 10^{123}, is therefore a logarithm of a probability, and therefore constitutes an information measure in bits (with numbers at this scale, we can ignore correction factors depending on the logarithmic base).

How big, in information-theoretic terms, is 10^{123} bits? That number of bits seems plenty big to account for all the information that might ever be required for biological systems. Human DNA is roughly 10^{10} bits of information (3 billion base pairs of nucleotides, each capable of carrying 2 bits of information). This number of bits is on the large side for organisms on planet earth, most of which are bacteria whose DNA is in the range of 10^6 to 10^7 bits of information. It is estimated that in the history of life on earth, at most 10^{40} organisms have ever existed.[46] That would yield an upper bound on the total information content of all the world's DNA at

[45]See [44]. Grant was a materialist, a freethinker, and an early mentor to Charles Darwin.

[46]See [46, p. 143]. Strictly speaking, Behe gives 10^{40} as an upper bound on "bacterial cells in the history of life on earth." But bacterial cells vastly outnumber other organisms. So 10^{40} is also a reasonable estimate for the total number of organisms that have ever existed on planet Earth.

$10^{10} \times 10^{40} = 10^{50}$ bits. That number is still minuscule compared to 10^{123}.

Of course, the case could be made that there is still much more information in life than can be captured in 10^{50} bits. But exhausting 10^{123} bits takes a lot of effort even at the scale of the entire universe. In any case, the point of these musings is to make clear that in terms of raw physics, there is plenty of information in the universe to accommodate the complexity and information-richness of life. Penrose's 10^{123} bits for the universe at large is actually on the low side. Seth Lloyd, a quantum computational theorist, estimates as many as 10^{240} bit operations at the scale of the entire universe throughout its duration, effectively increasing Penrose's 10^{123} bits by a factor of 10^{117}.[47]

From the vantage of physics, the universe therefore provides plenty of informational room for biology. It contains lots of information and can accommodate any degree of biological complexity we might encounter. In what sense, then, does biology face an information problem? The informational capacity of the universe provides a backdrop against which biology can flourish. Yet as a backdrop, it is a necessary but not a sufficient condition for life's origin and diversification. Physics, if you will, provides the informational infrastructure for biology. But the actual information embodied in biological systems needs more than infrastructure. How that information arises from that infrastructure needs itself to be explained.

Charles Darwin's claim to fame is that he proposed such an explanation. Darwin proposed his mechanism of natural selection acting on random variations to bring about the increases in information needed to build increasingly complex organisms ("monad to man"). Indeed, the idea that evolution must proceed from simplicity to complexity was central to Darwin's theory of evolution. Consider, for instance, the closing sentence of his *Origin of Species*: "There is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved" [47, p. 490]. The movement in Darwinian evolution is here clearly seen to be from simplicity ("so simple a beginning") to complexity ("forms most beautiful and most wonderful").

Richard Dawkins, as this age's preeminent disciple of Darwin, has fully developed the theme that evolution builds complexity from simplicity. In fact, Dawkins' approach to evolutionary biology and biological origins can be seen as predicated on denying the Law of Conservation of Information. If the Law of Conservation of Information applies to evolutionary biology, then evolution is a form of search and the emergence of information-rich biological systems expresses the redistribution of prior information rather than the creation of novel information. The Law of Conservation of Information thus explains subsequent informational complexity in terms of prior informational complexity where the prior informational complexity is at least as intensive as the subsequent informational complexity, implying that the biological information to be explained was in fact always there.

And yet, that is precisely what a self-contained non-teleological theory of evolution (such as Darwinism) cannot tolerate—namely, that the information was always there. To say that the information necessary for life was always there is as unacceptable on strictly materialistic evolutionary grounds as invoking an intelligent designer to explain that information. In that vein, consider the following quote by Richard Dawkins from *The Blind Watchmaker*:

> To explain the origin of the DNA/protein machine by invoking a supernatural Designer is to explain precisely nothing, for it leaves unexplained the origin of the Designer. You have to say something like "God was always there," and if you allow yourself that kind of lazy way out, you might as well just say "DNA was always there," or "Life was always there," and be done with it. [21, p. 141]

Given this remark, it is fair to say that Dawkins would also need to reject that "the information needed for life was always there." In fact, he makes this very point at the end of *The Blind Watchmaker* when he writes that we must not "assume the existence of the main thing we want to *explain*, namely organized complexity" [21, p. 316]. Organized complexity here is a synonym for the biological information that evolutionary theory needs to explain. Early in *The Blind Watchmaker*, Dawkins makes clear the connection between complexity and information:

> We were looking for a precise way to express what we mean when we refer to something as complicated... Complicated things have some quality, specifiable in advance, that is highly unlikely to have been acquired by random chance alone... My characterization of a complex object—statistically improbable in a direction that is specified not with hindsight—may seem idiosyncratic... [But] whatever we choose to call the quality of being statistically-improbable-in-a-direction-specified-without-hindsight, it is an important quality that needs a special effort of explanation. [21, p. 9,15]

[47] For a summary of Lloyd's work on the information-processing powers of the universe, see [19, pp. 180–181].

Note that Dawkins does not explicitly refer to this quality of being specifiable in advance and highly unlikely to have been acquired by random chance as information. But given the account of information presented earlier in this paper, that is exactly what it is. A rose by any other name is still a rose. Organized complexity is information.

So how do evolutionary biologists avoid presupposing the very information that they are called on to explain? According to Dawkins, only an evolutionary theory like Darwin's can legitimately explain organized complexity, and it does so by gradually accruing information so that more information comes out of the evolutionary process than is put into it. Echoing Darwin's conclusion in the *Origin*, Dawkins writes: "The one thing that makes evolution such a neat theory is that it explains how organized complexity can arise out of primeval simplicity" [21, p. 316].

Of all the things that Dawkins has ever written, this last quote provides as clear a window as any into why he regards Darwinism as indispensable to our understanding of the world. Committed as he is to materialism, Dawkins must explain how to get organized complexity from a cosmos that at some point in its past did not contain organized complexity (such as at the moment of the Big Bang). Only an evolutionary process that builds complexity from simplicity fits the bill. This is why Dawkins will also write, "Even if there were no actual evidence in favour of the Darwinian theory (there is, of course) we should still be justified in preferring it over all rival theories" [21, p. 287]. The parenthetical here—"there is, of course"—is a throw-away. On the assumption that informational complexity can only legitimately be explained in terms of prior informational simplicity, something like Darwinism must be true. Evidence becomes irrelevant.

The problem that Dawkins now faces, however, is not evidence but mathematics. In particular, the Law of Conservation of Information belies the productive informational regress that he needs for evolutionary theory to work. He needs the information for organized complexity to emerge from, as he puts it, a "primeval simplicity" that lacks such information. Such a movement from simplicity to complexity would constitute a productive informational regress. But as shown in Sections 6 and 11, evolution is a search for targets where both search and targets are defined in terms compatible with materialism. The Law of Conservation of Information thus applies. Consequently, Darwinian evolution, as an information-creation mechanism, is dead in the water. It simply cannot create the information that Darwinists need it to create.

As a postscript for readers who might still not be fully convinced about the power of conservation of information to overturn non-teleological evolutionary theories such as Darwinism, let me offer one further observation. Dawkins understands evolution as a movement from primeval simplicity to organized complexity and thus insists that biological complexity must always be explained in terms of prior simplicity. In fact, to read Dawkins, one would think that it is a prime directive of reason always to explain the complex in terms of the simple. But this view of rational explanation is clearly false. Shakespeare wrote *Macbeth*. It is explanatory to say that Shakespeare is the author of *Macbeth*. Yet who in their right minds would say that this is an invalid explanation because Shakespeare is more complex than *Macbeth* (surely he is)? Or who would say that we have not explained anything about *Macbeth* because we have not explained the origin of Shakespeare?

Dawkins, when pressed, realizes that explaining the complex in terms of the simple has its limits, even in the field of biological origins. This point became clear in an interview between Dawkins and Ben Stein for the 2008 documentary *Expelled: No Intelligence Allowed*. As Dawkins put it in that interview:

> It could be that at some earlier time, somewhere in the universe, a civilization evolved by probably some kind of Darwinian means to a very, very high level of technology and designed a form of life that they seeded onto, perhaps, this planet. Now, that is a possibility, and an intriguing possibility, and I suppose it is possible that you might find evidence for that. If you look at the details of biochemistry, molecular biology, you might find a signature of some sort of designer... But that higher intelligence would itself have had to have come about by some explicable or ultimately explicable process. It could not have just jumped into existence spontaneously.[48]

Dawkins here takes seriously the possibility that intelligent space aliens seeded life on planet Earth. But if one is willing to entertain that idea, why not entertain the idea that a higher intelligence independent of any simpler origin brought about life on Earth? Such an intelligence would be in a category of its own, not subject to explanation by appeal to more basic or primitive precursors. It would be irreducible and self-existent. In other words, it would be something like God. Of course,

[48] Dawkins in this interview portrays Darwinism as a Kantian synthetic a priori that is known independent of experience and so requires no evidence. Accordingly, Dawkins allows that life on Earth might have been designed by intelligent space aliens, but then immediately adds that those aliens would themselves ultimately need to be accounted for by an "explicable" (by which he means a Darwinian) process that did not start with intelligence. For the entirety of this remarkable interview, see https://www.youtube.com/watch?v=GlZtEjtlirc (last accessed January 11, 2025).

Dawkins is profoundly allergic to such a view. Yet it is worth pointing out that his resistance to God as a final resting place of explanation does not exempt him from also being committed to a final resting place of explanation—a place where all further explanation breaks down. The material world is for Dawkins fundamental and unexplained. He accepts it as a given. Yet the origin of the universe cries out for explanation, especially because it seems to have had a beginning (at the Big Bang) and because it seems to be winding down (on thermodynamic grounds).

Dawkins' requirement that the complex be explained in terms of the simple is a direct consequence of his materialistic worldview, not of objective science. Materialism is always a bottom-up affair, where particles arrange themselves into assemblies, which arrange themselves into assemblies of assemblies, and so on up. For materialism, higher degrees of complexity must thus properly be explained in terms of lower degrees of complexity. In general, materialism is reductionistic, committed to explaining wholes in terms of parts. But we all know that wholes can be more than the sum of their parts and that wholes are often needed to understand what the parts are doing. Wholes are more complicated than parts and yet can be explanatory of parts. There are Platonic, Aristotelian, Stoic, panpsychist, pantheistic, deistic, theistic, etc. worldviews that reject Dawkins' doctrine of explanation, which mistakenly requires the simple always to account for the complex.

A science open to the full range of possibilities and always willing to follow the evidence wherever it leads will look for the best explanations, irrespective of whether the simple explains the complex or vice versa.

13. APPLYING THE LAW OF CONSERVATION OF INFORMATION

In 2010, Robert Marks and I published "The Search for a Search: Measuring the Information Cost of Higher-Level Search," in which we proved our first conservation-of-information theorem for search [48]. In 2009, in the lead-up to proving this theorem, Marks and I published "Conservation of Information in Search: Measuring the Cost of Success" [13]. This article laid out the three key information measures involved in conservation of information for search—namely, endogenous information, exogenous information, and active/added information. These, as we saw at the end of Section 5, are equivalent to, respectively, the baseline probability, the improved probability, and the search-for-a-search probability (a logarithmic transformation connecting the information-theoretic and probabilistic versions).

The 2010 paper assumed that endogenous information was defined with respect to a uniform probability on the search space. Later, along with Winston Ewert, we proved a more general version of this theorem that omitted the uniform-probability assumption [49]. But this more general version, like the earlier, was formulated in terms of measure spaces of probabilities to represent search, thus making it more restrictive than the Law of Conservation of Information proved in this paper, which does not care how search is ultimately represented mathematically. These two earlier theorems are recounted in Appendix 2 of this paper (the last two of the four listed there). They are the ones on which all the subsequent published research on conservation of information for search has been based.

How much research have these initial papers by Marks, Ewert, and me inspired? While not overwhelming, peer-reviewed research stemming from these papers has also not been insubstantial. For instance, probabilist Daniel Andrés Díaz Pachón has headed a research project (involving collaborators such as Robert Marks) that uses conservation of information, and especially active information, as a tool for understanding a variety of scientific problems, everything from fine-tuning in cosmology to population genetics to hypothesis testing.[49] Bioinformaticist Steinar Thorvaldsen and statistician Ola Hössjer (mainly as a duo but also with other collaborators, including Díaz Pachón) have published a number of articles using active information to analyze the specified and irreducible complexity of biological systems, concluding that some of these systems "pose a serious challenge to a Darwinian account of evolution."[50] In general with this conservation-of-information-inspired research, the focus has been on using active information as a tool for understanding the input and output of information required to produce complex systems or to model complex tasks.

Unlike the research just described, which treats conservation of information pragmatically as an accounting tool for tracking active information, computer scientist George Montañez has extended the theoretical bounds of conservation of information. In conservation of information for search, the original focus was on a fixed target T and the information required to locate T by searching for a search. Montañez inverts this problem: Given a search ("search strategy" in his words), for which targets ("search problems" in his words) does the search perform well and for which does it perform poorly? Rather than searching for a search, Montañez fixes a search and examines how the search performs across different targets. In a result he calls "Conservation of Active Information of Expectations," the targets for which a search adds b bits of active information has a proportion, among all targets, of less than or equal to 2^{-b}, which corresponds to b bits of information [58].

In an intriguing parallelism, Montañez proves a conservation of information theorem for varying targets that

[49] On cosmological fine-tuning, see, for instance, [50–52]. See also [53, 54].

[50] Quoted from [55, p. 7]. See also [56] and [57].

exhibits the familiar p, q, and p/q probabilistic relations that characterize conservation of information in the search for single targets. Montañez uses this result to argue that machine learning works precisely because its algorithms thrive only in solving a limited number of problems, failing miserably on the vast majority of problems, leaving them unsolved. In addition, Montañez has proved some results that he has put under the rubric of "conservation of complex specified information" [59]. but these results are about generalizing specified complexity rather than placing informational limits on search. Finally, Montañez makes explicit the connections among search, biasing, and learning in a "conservation of bias" result that he and colleagues have proven and that can readily be reframed as a conservation of information result of the sort described in this paper [60, 61].

I have given here a thumbnail of how conservation of information in search has been applied within the intelligent design community in the years since Robert Marks and I published our first results in this area. Within that community, conservation of information has inspired active areas of research. Yet it might also be argued that researchers outside that community have been grappling with information-theoretic problems similar to those raised by conservation of information, proposing informational solutions that are at the very lease congruent with conservation of information. Even though they would not consider themselves to be intelligent design proponents, I mention here Michael Levin and his collaborators, whose application of information measures in their research leads them to conclude that Darwinian theory is fundamentally incomplete [62, 63]. Tom Dingjan and Anthony Futerman write about the fine-tuning of lipid bilayers in cell membranes, understanding the information needed to explain this fine-tuning in terms of what they call "compositional complexity" [64, 65]. Luciano Floridi deserves mention here as well for his extensive work on the philosophy of information [66–68].

I am gratified how earlier work on conservation of information by Marks and me has inspired a growing body of research. At the same time, I believe we've only scratched the surface of what conservation of information can do for science and engineering. By simplifying and generalizing our understanding of conservation of information, the Law of Conservation of Information as developed in this paper makes relations between informational input and output intuitively clear apart from performing precise mathematical calculations—by eyeballing rather than by calculating. Its applicability and versatility across STEM fields as a whole is thus greatly enhanced, allowing mathematical details to be worked out at leisure. As the following examples attest, the scope of the Law of Conservation of Information is very wide indeed. In fact, it might even be said that the real challenge now is to find areas where the law fails to apply.

1. Wheeler-DeWitt Equation in Cosmology. The Wheeler-DeWitt equation encapsulates the quantum state of the universe, producing probabilities for possible configurations of space and matter. Applying the Law of Conservation of Information, we see that any finely tuned outcome, such as a universe conducive to life, does not arise in the absence of prior information. Instead, the specific boundary conditions and symmetries encoded in the equation constrain the space of solutions, effectively channeling the relevant probabilities toward life-permitting configurations. This constraint highlights that the apparent order and complexity of the universe are not spontaneously generated but stem from pre-existing informational inputs, demonstrating that even fundamental physics inherits, rather than creates, information [69, ch. 17].

2. Biological Evolution and Functional Genomes. Evolutionary processes involve natural selection acting on genetic variations, but the improbability of forming functional genes or proteins by chance-based processes remains a challenge. The Law of Conservation of Information reveals that natural selection does not create new information from scratch but redistributes pre-existing information embedded in genetic variability, biochemical constraints, and environmental conditions. For instance, specific mutations may succeed due to fitness advantages, but the pathway to these mutations depends on already-existing molecular structures and selective pressures. Thus, the improbability of the chance-based formation of functional biological complexity can only be offset by improbable environmental constraints or an improbable evolutionary landscape [70, chs. 6–7].

3. Generative AI and Language Models. Generative AI, like GPT-4, produces coherent and contextually relevant text, but this ability depends entirely on vast, carefully curated training data and optimized neural network architectures. The Law of Conservation of Information shows that the improbability of generating fluent language output is inherited from the improbability of the pre-existing data and the supervised learning processes that fine-tune the model's weights. Without this painstakingly constructed foundation, the model would devolve into randomness. Every coherent sentence generated reflects the informational labor embedded in the training phase, ensuring that the high-quality output mirrors the improbable and highly structured input [71].

4. Weather Prediction Models. Accurate weather predictions rely on complex algorithms and an enormous influx of data from satellites, sensors, and historical records. The Law of Conservation of Information highlights that the improbability of correctly forecasting chaotic atmospheric conditions is mitigated by the pre-existing informational wealth of detailed observations and validated models of atmospheric physics. The predictive accuracy stems not from creating new information but from extracting patterns encoded in this extensive dataset. Without these high-quality inputs and carefully designed simulation models, weather prediction would be as unreliable as a random guess, underscoring the need for these models to inherit informational precision [72].

5. Genetic Algorithms in Optimization Problems. Genetic algorithms solve optimization problems by simulating evolutionary processes, iteratively refining solutions over generations. Their success depends on fitness functions and constraints explicitly designed to evaluate and guide each generation toward better solutions. According to the Law of Conservation of Information, the improbability of finding an optimal solution is transferred to the fitness function's embedded knowledge about the problem space. These algorithms rely on pre-existing informational inputs to guide their search, meaning that their apparent efficiency derives from leveraging carefully encoded domain expertise, not from spontaneously generating problem-solving information [73].

6. Speech Recognition Systems. Speech recognition software like Siri or Alexa transcribes spoken words into text with remarkable accuracy, but this capability hinges on extensive training datasets and linguistic models. The Law of Conservation of Information reveals that the improbability of correctly decoding spoken language is offset by the curated training data, which captures phonetic, grammatical, and contextual nuances. The software does not create the ability to understand speech; it inherits it from the design of its neural networks and the exacting labor that pre-annotated the training data. Without this foundational information, speech recognition would fail to distinguish between even basic words [74].

7. Cognitive Behavioral Therapy in Psychology. Cognitive Behavioral Therapy (CBT) helps individuals overcome negative thought patterns, but its success depends heavily on structured therapeutic models and the prior experiences of both patient and therapist. According to the Law of Conservation of Information, the improbability of finding an effective strategy for behavioral change is counterbalanced by pre-existing psychological theories, therapeutic frameworks, and the personal insights brought to therapy sessions. CBT's effectiveness reflects the embedded information in these structured approaches, not an unstructured creation of solutions. The method's power comes from skillfully applying this informational reservoir to individual challenges, showcasing the informational transfer of therapeutic insight [75].

8. Market Predictions in Economics. Predicting stock market trends is notoriously difficult due to the chaotic interplay of countless variables. Tools like econometric models and historical data analysis attempt to improve the accuracy of these predictions. The Law of Conservation of Information applies here: the improbability of accurately forecasting a market outcome is reduced only through pre-existing knowledge about market behaviors, historical trends, and model constraints. Even advanced methods, like machine learning algorithms for market analysis, inherit their predictive power from the quality and depth of their training data. This ensures that any improvement in prediction accuracy reflects the importation of prior economic information [76].

9. Legal Precedents and Judicial Rulings. Judicial decisions often depend on applying legal precedents to new cases, effectively making rulings more predictable and fair. The Law of Conservation of Information applies because the improbability of deriving just rulings in complex cases is reduced by pre-existing legal frameworks, case law, and statutory guidelines. Judges and attorneys use these sources to structure arguments and decisions. Without these foundational precedents and principles, the judicial process would devolve into arbitrary decisions, highlighting how rulings inherit informational constraints from the accumulated history of legal thought [77].

10. Drug Discovery in Pharmaceuticals. Developing a new drug is akin to searching for a needle in a haystack, with countless potential molecular combinations to explore. The success of this search depends on pre-existing biochemical knowledge, such as protein structures, disease mechanisms, and pharmacodynamics. The Law of Conservation of Information ensures that the improbability of discovering an effective drug is transferred to the immense informational input required to guide the search. Computational models and empirical data from prior research allow scientists to focus their efforts, demonstrating that the breakthroughs in drug discovery reflect inherited, not spontaneously generated, information [78].

14. CONCLUSION: THE ULTIMATE SOURCE OF INFORMATION

The Law of Conservation of Information shows that information cannot emerge spontaneously in sufficient quantity to resolve needle-in-a-haystack problems. These are problems whose successful resolution via search is vastly improbable. Instead, information that facilitates search must be derived from prior inputted information, which must be at least as substantial as outputted information. This fundamental relationship between information inputted and information outputted implies a conundrum for any attempt to explain the ultimate origins of information. Tracing information back to prior sources reveals an ever-intensifying challenge, as each higher-level round of search to explain information demands an account of still greater prior informational input. Without some final resting place of explanation, we confront a regress that cannot resolve itself. A question then becomes inescapable: What is the ultimate source of this information, and how does it fit within our understanding of reality?

Three main options exist for answering this question: (1) front-loading, in which all the information required for subsequent developments is encoded at the universe's outset; (2) brute chance, in which the sheer number of opportunities that reality makes available, as with a multiverse, overwhelms any improbabilities; (3) irreducible intelligence, where the source of information is beyond the constraints of physical processes. Each of these options has profound implications for our understanding of the cosmos and our place in it.[51] Here, I will evaluate each in light of the Law of Conservation of Information and argue that the influence of an irreducible intelligence best explains the ultimate source of information.

The first option, front-loading, suggests that all the necessary information for the universe's unfolding—from the formation of galaxies to the emergence of complex life—was embedded within the universe's initial conditions. More precisely, the manifestation of all such information as the universe unfolds traces back to an initial deposit of information, which is the information that was front-loaded. For this front-loaded information, there are no prior informational precursors. Within the front-loading perspective, this front-loaded information becomes the source of all information downstream from it, especially information responsible for successful targeted search.

Interestingly, because this front-loaded information is embedded in the universe from the start, its causal antecedents, such as they may be, must reside outside the universe (otherwise the information would not be front-loaded but subsequent to other information in the universe causally responsible for it). Accordingly, any probabilities associated with this front-loaded information will be impossible to ground empirically but must reside within a broader framework of probabilities by which universes come into existence. To be empirically grounded, probabilities must operate within an already existing backdrop of physical causation (such as the universe or some portion of it). Front-loaded information, by contrast, must be logically antecedent to any such backdrop. Any probabilities associated with this front-loaded information will therefore have to be assigned on theoretical rather than empirical grounds.

Proponents of front-loading point to such remarkable findings as the fine-tuning of physical constants, which is needed for life to exist at all, or the mathematical elegance of natural laws, which, as Eugene Wigner put it, "is a wonderful gift which we neither understand nor deserve" [79, p. 14]. From a front-loading perspective, the Big Bang was not merely a chaotic explosion but an event laden with potential, akin to a seed containing all the instructions to grow into a mature organism. Front-loading provides an answer to conservation of information's regress problem (Section 12) by simply stipulating an end to the regress at the point of front-loading. But that raises troubling questions of its own. Why should we have such a front-loaded universe at all? Where, in the primordial plasma of the early universe, is the information necessary to produce, say, a minimally complex cell—a feat that defies even our most advanced engineering capabilities?

Front-loading shifts the explanatory burden back to the universe's earliest conditions, such as the Big Bang, but it does not ultimately resolve the question of information origins. What mechanism or cause encoded this information into the universe in the first place? Front-loading, though intriguing, risks becoming an unsupported article of faith unless it can account for how such a densely packed repository of primordial information came to be. To date, no satisfying naturalistic account of this primordial information is on offer.

Philosopher Holmes Rolston, independently of the Law of Conservation of Information, captures its lesson for front-loading. He persuasively critiques the idea that all biological information was front-loaded into the universe at its inception. At the same time, as a non-materialist, he avoids attributing this information to a non-teleological evolutionary process, whether cosmic, chemical, or biological. In an insightful and detailed passage, he writes:

> There are no humans invisibly present (as an acorn secretly contains an oak) in the primitive eukaryotes, to unfold in a lawlike or programmatic way... On Earth, there really is not anything in rocks that suggests the possibility of *Homo sapiens*, much less the American Civil War, or the World Wide Web, and to say that all these possibilities are lurking there, even though nothing we know about rocks or carbon atoms, or electrons and protons suggests this is simply to let possibilities float in from nowhere...
>
> The information (in DNA) is interlocked with an information producer-processor (the organism) that can transcribe, incarnate, metabolize, and reproduce it. All such information once upon a time did not exist but came into place; this is the locus of creativity. Nevertheless, on Earth, there is this result during evolutionary history. The result involves significant achievements in cybernetic creativity, essentially incremental gains in information that have been conserved and elaborated over evolutionary history. The know-how, so to speak, to make salt is already in the sodium and chlorine, but the know-how to make hemoglobin molecules and lemurs is not secretly coded in the carbon, hy-

[51]The best treatment of these options is [69].

> drogen, and nitrogen... Can one claim that what did actually manage to happen must always have been either probably probable, or, minimally, improbably possible all along the way?
>
> Push this to extremes, as one must do, if one claims that all the possibilities are always there, latent in the dust, latent in the quarks. Such a claim becomes pretty much an act of speculative faith, not in present actualities, since one knows that these events took place, but in past probabilities always being omnipresent... Unbounded possibilities that one posits ad hoc to whatever one finds has in fact taken place—possibilities of any kind and amount desired in one's metaphysical enthusiasm—can hardly be said to be a scientific hypothesis. This is hardly even a faith claim with sufficient warrant. It is certainly equally credible and more plausible, and no less scientific to hold that new possibility spaces open up en route. [80, pp. 352–353, 357]

The second option, brute chance, attempts to sidestep the improbability of our universe's information-rich structure by positing some entity, such as a multiverse, that vastly expands the number of opportunities for chance to operate. Given an infinite or near-infinite number of universes spanning the full range of boundary conditions and allowing unlimited do-overs, the emergence of a universe capable of supporting complex life—though astronomically improbable in isolation—becomes inevitable across the multiverse's vast probabilistic landscape. This approach effectively reduces the question of information origins to a statistical inevitability, bypassing the need for further causal explanation. In the notation of the Law of Conservation of Information, $P(E) = p$ is now no longer so small as to make attaining E require heroic input from external information. In this way, all needle-in-a-haystack problems become solvable by brute chance.

Yet, this solution is fraught with its own challenges. Notably, it lacks empirical support. While the multiverse (or similar devices that vastly inflate probabilistic resources, whether locally or globally) remains a speculative hypothesis in cosmology, it offers no mechanism for observing or interacting with other universes, making it untestable and unfalsifiable. Also, it sidesteps rather than addresses the problem of information origins. Even if a multiverse exists, why should it produce informationally rich universes, such as exhibit finely tuned parameters? Indeed, what prevents all the universes within a multiverse from being simplistic and boring? Even a multiverse requires a principle of plentitude to ensure that interesting universes are included within it.[52] There is also the practical challenge that positing brute chance as the ultimate source of information abandons science's responsibility to offer insightful explanations. When a bridge collapses, we do not invoke brute chance; we seek specific causes. To do otherwise undermines the very enterprise of science.[53]

The third option, irreducible intelligence, identifies the ultimate source of information with an active, purposeful agent capable of introducing novel information into the cosmos. Such an intelligence would be irreducible to physical processes, transcending the probabilistic constraints that bind physical systems. It would not merely shuffle or repackage preexisting information but rather originate new information in creative acts. Such an irreducible intelligence is not as radical as it might initially seem. In quantum mechanics, for instance, the role of the observer highlights the porous boundary between mind and matter, suggesting that consciousness interacts with physical systems in ways that could defy reductionist explanations. Moreover, intelligence as a causal principle aligns with our everyday experience: whenever we encounter complex specified information—such as a book, a blueprint, or a computer program—we attribute it to a mind, not to chance and necessity.[54] Indeed, humans are sources of information that readily find targets effectively impossible to find by accident.[55]

But how do humans accomplish this feat? And how, ultimately, would a nonmaterial designer behind the universe accomplish this feat of originating novel information? The subtitle of this paper is "Search Processes Only Redistribute Existing Information." We live in an age of "process" where everything that happens is supposed to happen via a sequence of steps, each dependent on the one preceding it. It sees everything that happens

[52] Belief in a multiverse, if it is to go anywhere, needs to be combined with a belief that the multiverse includes enough interesting worlds, perhaps all physically possible worlds, perhaps all logically possible worlds. That is what I mean by a principle of plenitude—there need to be enough interesting worlds. Philosopher David Lewis, with his modal realism, took the principle of plenitude as far as it could be taken by regarding all possible worlds as equally real to our own. For his account of modal realism, see [81].

[53] For the multiverse and its inflation of probabilistic resources, see [82, ch. 13].

[54] [19] argues that complex specified information is the key identifier of intelligence. Douglas Robertson, in a related vein, connects information creation to intelligence. In a prescient article that examined a conservation law for algorithmic information theory [AIT], he suggested that the defining feature of intelligence is its ability to create information that is beyond the remit of computational processes, and thus by implication beyond the remit of chance and necessity more generally. Identifying intelligence with free will, he wrote, "Free will appears to create new information in precisely the manner that is forbidden to mathematics and to computers by AIT" [83, p. 26].

[55] I owe this apt way of stating the matter to Douglas Axe.

as path dependent. This view is deeply evolutionary. It is also delusional. The merest reflection on human experience and mental life shows that much of what we think, say, and do is discontinuous, underdetermined by informational precursors, which may be necessary but by no means sufficient to explain informational outputs that need to be explained.

Humans have always needed external information that cannot be reduced to search processes or any other processes dependent strictly on prior information. Consider the need for oracles among the ancients. Homer, for instance, invoked the muse at the start of the *Iliad* to inspire his epic poem. Not just artists but also scientists require such inspiration. Consider Henri Poincaré (1854–1912), widely regarded the greatest mathematician of his age, surpassing even David Hilbert, as evidenced by Poincaré being awarded the Bolyai Prize in 1905 ahead of Hilbert.[56] Here is how Poincaré described one of his outstanding mathematical discoveries, which involved Fuchsian functions:

> Having reached Coutances, we entered an omnibus to go some place or other. At the moment when I put my foot on the step the idea came to me, without anything in my former thoughts seeming to have paved the way for it, that the transformations I had used to define the Fuchsian functions were identical with those of non-Euclidean geometry. I did not verify the idea; I should not have had time, as, upon taking my seat in the omnibus, I went on with a conversation already commenced, but I felt a perfect certainty. On my return to Caen, for conscience' sake I verified the result at my leisure.[57]

A process reductionist—someone committed to reducing all human creativity to path dependent processes—would contend that Poincaré's mind here was merely operating as a process. In our day, process reductionism is usually conceived as computational reductionism. Accordingly, the crucial computations needed to resolve Poincaré's theorem would have been going on in the background somewhere in his brain and then just happened to percolate into his consciousness once the underlying computations had run their course. But the actual experience and self-understanding of thinkers like Poincaré, in accounting for their bursts of creativity, gives no evidence of issuing from any underlying process—neurophysiological, computational, or otherwise.

Humanists have always—and rightly—rejected a reductionist view of human creativity and intelligence. For instance, Joseph Campbell, best known for his work on comparative mythology, offered this rejoinder to computational reductionism: "Technology is not going to save us. Our computers, our tools, our machines are not enough. We have to rely on our intuition, our true being" [84]. In mental life, intuition and discursive reason play distinct but complementary roles. Intuition—also called insight, instinct, or immediate apprehension—is the mind's ability to grasp truths, patterns, or solutions instantly, without the need for step-by-step reasoning. It operates holistically and often subconsciously, drawing on experience, implicit knowledge, or even an inexplicable sense of certainty. By contrast, discursive reason—also known as logical reasoning, analytical thinking, or deduction—proceeds methodically, moving from premises to conclusions, and more generally from inputs to outputs via a rules-based path-dependent process.

Both aspects of the mind are necessary for finite rational agents like ourselves. Great intellectual achievements, from scientific discoveries to philosophical breakthroughs to artistic triumphs, typically emerge from their dynamic interplay—where flashes of intuition ignite the creative discovery, and then mechanical forms of reasoning demonstrate its soundness. While intuition offers swift, often brilliant leaps of understanding, it can be mistaken, requiring the rigor of discursive reasoning to test, clarify, and justify its insights. Conversely, discursive reasoning, though reliable and systematic, can be slow and stodgy, limited by available premises, and at times incapable of grasping truths that transcend formal logic. The two faculties work best in concert: intuition generates novel insights, while discursive reason refines and verifies them.

This duality of the mind in which the intuitive and the discursive play off each other is reflected in the role of the right and left hemispheres of our brains. The two hemispheres make sense of reality in fundamentally different ways, and the balance between them shapes human culture, intelligence, and creativity. The left hemisphere is analytic, detail-oriented, and focused on abstraction and categorization. It excels in breaking down reality into discrete parts, making it essential for language, logic, and technical problem-solving. However, it tends to oversimplify and fragment reality, leading to a mechanistic and reductionist worldview. In contrast, the right hemisphere engages with the world holistically, perceiving relationships, ambiguity, and lived experience. It is attuned to metaphor, emotion, and the interconnectedness of reality, making it central to creativity, intuition, and wisdom.[58]

[56] See https://en.wikipedia.org/wiki/Bolyai_Prize.

[57] Quoted from Eric-Jan Wagenmakers, "Henri Poincaré: Unconscious Thought Theory Avant la Lettre," available at https://www.bayesianspectacles.org/henri-poincare-unconscious-thought-theory-avant-la-lettre (last accessed March 4, 2025).

[58] In this paragraph and in the subsequent discussion about right and left hemispheres, it may seem that I am ascribing mental powers to these parts of the physical brain, and that this manner of description is in tension the thesis I am arguing that such mental powers constitute an irreducible intelligence. But this is not the case. It must be recognized that physicalism need not be the same thing as materialism. Even though I don't subscribe to

According to neuroscientist Iain McGilchrist, true intelligence emerges from a dynamic integration of both hemispheres, where the right hemisphere provides a broad, contextual understanding, and the left hemisphere corrects, refines, and articulates insights. McGilchrist contends that modern culture has become dominated by the left hemisphere's mode of thinking, prioritizing rigid logic, efficiency, and control (process) at the expense of depth, insight, and creativity (intuition). This imbalance, he suggests, has led to a disconnected view of reality along with a concomitant loss meaning and wisdom. Creativity flourishes when the right hemisphere leads the way, allowing for insight and novel connections, while the left hemisphere helps structure and articulate those insights. A flourishing mind, society, and culture, McGilchrist argues, require the right hemisphere to guide the left, rather than allowing the left hemisphere's fragmented, mechanistic perspective to dominate our understanding of the world [85, 86].

Consistent with the operation of our right hemisphere, an irreducible intelligence would be capable of creating information *de novo*, intuiting solutions to what otherwise would be probabilistically intractable needle-in-a-haystack problems. Indeed, an irreducible intelligence could be defined as the capacity to overcome improbabilities by intuiting truths directly without having to engage in a search process. Process reductionists, who insist that all intelligence must reduce to processes, will reject such an irreducible intelligence. But then they must contend with human experience, which, as in the example of Henri Poincaré, provides evidence for intuition solving problems beyond the reach of processes. Indeed, we lack any process account whatsoever for Poincaré's insight into Fuchsian functions. Poincaré's case, though remarkable, is hardly unique. The human mind has a long history of producing flashes of insight that remain thoroughly unaccounted for in terms of any process.

In pooh-poohing the idea of an irreducible intelligence, process reductionists must above all contend with the Law of Conservation of Information. The Law of Conservation of Information establishes that information needed for successful search cannot bootstrap itself into existence. If the universe's information-rich structure cannot be plausibly explained by front-loading or brute chance, then the introduction of information by an irreducible intelligence, acting without being reducible to any process, becomes a viable alternative. This intelligence would not be a product of the universe but its source, existing beyond it, yet capable of interacting with and palpably influencing it by inputting newly created information. Such an intelligence avoids the information regress problem by grounding information in a singular, transcendent source that creates the information without needing to derive it from a search process.

Most people would equate such an irreducible intelligence with a deity (whether theistic, deistic, panentheistic, or pantheistic). But other explanatory options exist. The simulation hypothesis, for instance, characterizes the universe and its informational structure as the result of an advanced computational simulation [87, ch. 9]. Such a simulation would be orchestrated by entities that might in some sense be material or physical. And yet the entities responsible for such a computational simulation would be external to our universe. This scenario reframes the intelligence responsible for information not as transcending all physicality but as existing within a higher-order physical reality. In any case, such an intelligence would then be irreducible to the physical reality that we inhabit.

Obviously, such a simulation would raise the prospect of the simulation itself residing in a still higher-order simulation, and so on, issuing in a potentially infinite regress. Thus, our world could be a simulation created by hackers who are themselves in a simulation created by higher-level hackers, and so on indefinitely. Such a regress raises doubts about the epistemological soundness of the simulation hypothesis. But the bigger challenge facing the simulation hypothesis is that it presupposes a computational reductionism. The simulation hypothesis makes all of reality computational. Yet we have no convincing evidence to think that all of reality can be digitized (especially in light of conservation of information). For instance, in our ordinary experience, the organic cannot be reduced to the digital. In particular, we have zero evidence that turning bits on and off according to some algorithm equates with human consciousness.[59]

physicalism, I want to leave room for a non-reductive physicalism in which mental powers, and hence minds exhibiting irreducible intelligence, can be a property of a physical system such as a brain. Such mental powers would then not be reducible to material processes characterized in terms of the motions and modifications of material particles operating according to non-telic material forces. Physicalism can allow that there is more to the physical than pure materiality, with mind being an emergent property of the physical. If this sounds like nature mysticism, it probably is. My own predilection is that the brain, and in particular its right and left hemispheres, are physical systems that act in ways not reducible to materialism (thereby at odds with computational reductionism), but that also tap into a non-physical dimension of intelligence. In any case, ascribing various mental powers or activities to the brain or portions of it in this discussion should not be interpreted as a concession to materialism or a reduction to materialistic processes.

[59] To expand on this point, the prime weakness of the simulation hypothesis is that it is wedded to a computational view of intelligence, with all the problems that raises by reducing mind to mechanism. Indeed, the simulation hypothesis delivers an irreducible intelligence only in the sense that the simulators are not part of the simulation. At the same time, this hypothesis cannot help but deliver an intelligence reducible to computational mechanism since such mechanism is responsible for the simulation. Mechanism in all its guises has consistently run aground on the mind-body problem. In particular, it offers no place for a robust

Traditional theism blocks computational reductionism by denying that God, the source of reality, is a computer. God is not a material thing. God does not consist of parts. God is not a concatenation of bits. God is not a Turing machine. God is a free spirit, unconstrained by any created thing. And God alone is uncreated. Moreover, traditional theism blocks the regress implicit in the simulation hypothesis by making God the ultimate resting place of explanation, affirming God's self-existence (the theological term for this is "aseity") and thereby heading off the question of who created God.[60]

The irreducible intelligence option, which includes but is not coextensive with traditional theism, accords with the widespread intuition that information reflects intentionality. Just as a coded message implies a sender, the intricate informational architecture of the universe implies a source capable of intention. Critics may argue that invoking irreducible intelligence introduces metaphysical assumptions into what should be a purely scientific inquiry. Yet, the alternatives also rest on metaphysical assumptions: front-loading assumes unexplained initial conditions, and brute chance assumes an unobservable multiverse or some other device that is equally unobservable.

Strictly speaking, the three options considered here (front-loading, brute chance, and irreducible intelligence) are neither mutually exclusive nor exhaustive. Other options exist, such as a cyclical or eternal universe that can explore all of physical possibility space. But with respect to the Law of Conservation of Information, this option is effectively equivalent to the brute-chance option. Emergentism and panpsychism might also be cited as distinct options, but neither can account for nature executing improbable searches except by postulating that nature has this as an in-built capacity. And how is that different from such a capacity being front-loaded or resulting from brute chance in a multiverse?

Among the three options considered, there is also overlap, though they are sufficiently different to deserve being distinguished. Front-loading naturally calls to mind a front-loader, and so this option could be seen as falling under the irreducible intelligence option. Yet the point of the front-loading option, as described here, is not to draw a design inference but simply to note that the Law of Conservation of Information traces a universe with a beginning to a starting point of maximal information. Such maximal information at the outset is, within this option, a surd—an axiom that cannot be further analyzed. Of course, if front-loading is then in turn ascribed to an irreducible intelligence, such as a creator God, then we are dealing no longer with front-loading per se but instead with the irreducible intelligence option (albeit more deistic than theistic in flavor). Hybrids among these options exist as well. A multiverse could include a universe that is front-loaded. An irreducible intelligence might create a multiverse. Etc.

In conclusion, the Law of Conservation of Information reveals that information cannot emerge spontaneously or regress indefinitely without a foundational source. While front-loading and brute chance attempt to address this challenge, they do not so much explain as explain away information origins. An irreducible intelligence, transcending our physical world, yet capable of influencing it, offers a compelling alternative. Besides grounding information origins in intelligence, it also situates humanity in a universe imbued with intentionality. Recognizing this intelligence opens the door to a wider understanding of reality, one in which information is not merely a byproduct of chance and necessity but a reflection of meaning and purpose.

APPENDIX 1: A BRIEF HISTORY OF CONSERVATION OF INFORMATION

The term "conservation of information" appears in both physics and computer science, but its meaning differs in these fields. In physics, it applies to reversible processes where one physical state can be recovered without loss from another, ensuring information is strictly conserved between states. In computer science, it applies to search processes where information output cannot exceed prior information input, with information being strictly conserved when the quantity of information outputted equals the quantity inputted, and unconserved otherwise.

These uses of the term "conservation of information" intersect when physical processes are involved in search. But even then, they remain quite distinct. In physics, "conservation" always implies strict constancy, as with energy in a closed system, where total energy remains

---

conception of consciousness. Leibniz understood this weakness of reducing mind to mechanism clearly in the early 1700s, putting it as follows in his *Monadology*:

> We are obliged to admit that *perception* and that which depends on it *cannot be explained mechanically*, that is, by means of shapes and motions. And if we suppose that there were a machine whose structure makes it think, feel, and have perception, we could imagine it increased in size while keeping the same proportions, so that one could enter it as one does with a mill. If we were then to go around inside it, we would see only parts pushing one another, and never anything which would explain a perception. This must therefore be sought in the simple substance, and not in the compound or machine.

See [88, p. 17]. I challenge computational reductionism of mind in "Artificial General Intelligence as an Idol for Destruction," available on my blog at https://billdembski.com/artificial-intelligence/artificial-general-intelligence-idol-for-destruction—see especially Section 9 in that article on the difference between organisms and mechanisms (last accessed December 12, 2024). For a thorough deconstruction of computational reductionism and materialist accounts of mind more generally, see [89].

[60]For aseity, see [90, pp. 36, 40, 684].

constant though its form may change (e.g., potential, kinetic, or chemical energy). By contrast, in computer science, conservation represents an ideal limit or upper bound that search processes may achieve but often do not, reflecting an asymmetry between input and output information.

Conservation of information in physics arises in a variety of situations. Stephen Hawking, for instance, memorably proposed that information represented in matter may be irretrievably lost once sucked into a black hole that then evaporates in what has come to be called Hawking radiation. Under these circumstances, Hawking saw conservation of information as violated. Leonard Susskind, by contrast, proposed a holographic shifting of information from higher-dimensional spaces to lower-dimensional boundaries along with the idea of black hole complementarity. Accordingly, he theorized that information encoded on the event horizon could be recovered from Hawking radiation in highly scrambled form, preserving information rather than destroying it. In this way, Susskind saw even black holes as obeying conservation of information.[61]

In classical thermodynamics, conservation of information is tied to the second law of thermodynamics and the concept of entropy. While entropy has an inherent tendency to increase, suggesting a loss of order or information in macroscopic systems, statistical mechanics treats the underlying microstates of a system as unfolding deterministically and therefore keeping information intact. Accordingly, conservation of information becomes a consequence of ergodic theory, in which all states of a suitably configured dynamical system will recur. In statistical mechanics, the relevant result from ergodic theory is Liouville's theorem, which states that the phase-space density of a closed system remains constant over time, implying that information about initial conditions is conserved even if it appears jumbled at a macroscopic level [92, sec. 2.2].

Other examples of conservation of information in physics can readily be found. In quantum mechanics, the unitary evolution of quantum states as governed by the Schrödinger equation preserves the total information encoded in a quantum state. This conservation of quantum information was also underscored in the more recently proven no-hiding theorem [93]. Classical physics, which is deterministic and whose equations are time-reversible, implies that information about a system at any one point in time can, in principle, be reconstructed from any other point in time, thereby conserving information. Laplace even argued that conservation of information in this sense (though without using this terminology) applied across the entire universe. Thus, over 200 years ago in his *Celestial Mechanics*, he argued that from Newton's physics it followed that knowing the positions and momenta of all particles at a given moment in time could be used to predict and retrodict every aspect of the universe's unfolding down to the tiniest details.[62]

I recount these examples of conservation of information in physics not only to indicate that conservation of information is a respectable notion within physics but also to distinguish its role in physics from its role in search. In physics, conservation of information refers to the precise preservation of the same quantity of information through various physical transformations, with any one instance of information being recoverable from any other. Yet when it comes to search, conservation of information asserts that information outputs never exceed information inputs. Conservation of information in search thus says you cannot get something from nothing, but it allows that starting with something you might end up with less of that thing or even nothing.

Conservation of information in search is more practical than conservation of information in physics. From the vantage of conservation of information in physics, the destruction of the Library of Alexandria by fire did nothing to destroy the information in its volumes—all the information still exists and is even recoverable if we could but reverse in time the physical processes at play since the library's destruction. A super-physicist could thus in principle reconstruct the entire Library of Alexandria. Conservation of information in search,

[61] [91, ch. 9]. The title of this chapter is, tellingly, "The Puzzle of Information Conservation in Black Hole Environments."

[62] As Laplace put it:

> We ought then to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow. Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it—an intelligence sufficiently vast to submit these data to analysis—it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes. The human mind offers, in the perfection which it has been able to give to astronomy, a feeble idea of this intelligence. Its discoveries in mechanics and geometry added to that of universal gravity, have enabled it to comprehend in the same analytical expressions the past and future states of the systems of the world. Applying the same method to some other objects of its knowledge, it has succeeded in referring to general laws observed phenomena and in foreseeing those which given circumstances ought to produce. All these efforts in the search for truth tend to lead it back continually to the vast intelligence which we have just mentioned, but from which it will always remain infinitely removed. This tendency, peculiar to the human race, is that which renders it superior to animals; and their progress in this respect distinguishes nations and ages and constitutes their true glory.

Quoted from [94, pp. 17–18].

however, recognizes "frictional forces" that lead to inefficiencies in information transfer, treating information loss as the norm and full preservation as a seldom-attained ideal. From the vantage of conservation of information in search, the Library of Alexandria is indeed lost, a view shared by archeologists and historians.

The actual term "conservation of information" makes more than a passing appearance first in the writings of biologist and Nobel laureate Peter Medawar. There, in a slender volume titled *The Limits of Science*, published in 1984, he includes a Section with the same title as this paper: "The Law of Conservation of Information" [95, pp. 78–82]. Medawar formulated this law as follows: "No process of logical reasoning—no mere act of mind or computer-programmable operation—can enlarge the information content of the axioms and premises or observation statements from which it proceeds" [95, p. 79]. Rather than defend this law, Medawar treated it as self-evident: "I attempt no demonstration of the validity of this law other than to challenge anyone to find an exception to it—to find a logical operation that will add to the information content of any utterance whatsoever" [95, p. 79].

Medawar explains in the preface to *The Limits of Science* that he drew inspiration for this law and the terminology of "conservation of information" from physicist H. A. Rowland, who in 1899 presented a paper to the American Physical Society titled "The Highest Aim of the Physicist." In that paper, Rowland formulated "the law of the conservation of knowledge" as follows: "A mathematical investigation always obeys *the law of the conservation of knowledge*: we never get out more from it than we put in. The knowledge may be changed in form, it may be clearer and more exactly stated, but the total amount of the knowledge of nature given out by the investigation is the same as we started with" [96, emphasis added]. Medawar's law and Rowland's law are equivalent.

For Medawar and Rowland, the term "conservation" is supposed to capture the idea that for logical, algorithmic, or mathematical processes, outputs of knowledge or information cannot exceed prior inputs of the same. But others have expressed this idea as well, with or without using the term "conservation." In a book on science and information published in 1962, the French physicist Léon Brillouin made exactly the same point:

> It has been often hinted that a computing machine actually manufactures new information. This is not really the case. Machines are able to process information; they take the raw material and they give out a finished product, but the total amount of information has not been increased. In the best ideal circumstances, information may be kept constant during the computing, but under normal conditions there will be some loss and the final information will be smaller than the input. [97, p. 267]

And what is the ideal circumstance in which no information is lost? According to Brillouin,

> If the machine [for computing firing tables of projectiles] works perfectly it must be reversible: given the impact [of the projectiles] it may compute backward and obtain the initial conditions. In a similar way, a translator may translate back into Japanese, and should recover the original paper or its equivalent. Ideal operating conditions correspond to reversibility, hence to *conservation of information*. [97, p. 268, emphasis added]

This passing reference to conservation of information is the earliest I have been able to find in the literature. Brillouin sums up the matter thus: "The [computing] machine does not create any new information, but it performs a very valuable transformation of known information" [97, p. 269].

One way information theorists justify that algorithmic processes do not create but only redistribute existing information is through what is called the *data processing inequality*.[63] This result from information theory states that information cannot be increased by simply processing data. Starting with some original information, and then passing it through an algorithmic process, the total resulting information can never exceed what was there at the start. The process may lose or distort some of that information, but it cannot add anything new. Consistent with conservation of information, the data processing inequality underscores that no computational or algorithmic device can magically generate new information. Rather, such machines can only work with what they are initially given, so that in the course of their operation they may actually lose some of the original information.

The idea that machines can at best output what they first received as input goes well back to the first half of the nineteenth century to Charles Babbage's Analytical Engine. In its design, this was first full-fledged digital computer, though given the materials of the time Babbage was unable to make it fully functional. Babbage's

[63] See [5, pp. 34–35]. The data processing inequality is stated in terms of mutual information. Specifically, as information passes from X to Y to Z ($X \to Y \to Z$) and if X and Z are conditionally independent given Y (which is to say, $X \to Y \to Z$ forms a Markov chain), their mutual information satisfies $I(X;Y) \geq I(X;Z)$. This inequality implies that any transformation or processing of X by Y into Z cannot generate new information about X; it can only preserve or lose information. The data processing inequality reflects the inherent limitations of information flow in computational or algorithmic systems, ensuring that no operation can create information that was not already present in the input data.

Analytical Engine was designed to be a mechanical digital computer, not an electronic digital computer as we have in our day and that has proven itself to be workable in practice. Yet Babbage and his colleague Ada Lovelace understood the theoretical implications of their computing machine, even if they were never able to get a full working model. As Lovelace put it in her 1843 notes about Luigi Menabrea's memoir on Babbage's Analytical Engine,

> The Analytical Engine has no pretensions whatever to *originate* anything. It can do whatever we *know how to order it* to perform. It can *follow* analysis; but it has no power of *anticipating* any analytical relations or truths. Its province is to assist us in making *available* what we are already acquainted with.[64]

The idiom here is slightly different, but the point is exactly the one made by Medawar and Brillouin.

A few years earlier, in 1836, Edgar Allen Poe likewise made the same point, and again in addressing Babbage's Analytical Engine. Poe, in an essay titled "Maelzel's Chess Player," investigated the claim that this supposed chess playing machine was an automaton that could play chess without human intervention. As he saw it, Maelzel's chess player was a subterfuge: it was made to appear as though a machine was playing chess all by itself even though a human hidden inside it was in fact playing chess. Thus, while Maelzel would make a show of letting the audience see the interior of his chess player, in fact the full workings of the interior were always occluded, allowing a small human to lurk and scurry about inside the so-called machine. As Poe put it,

> The man within is now at liberty to move about. He gets up into the body of the Turk [i.e., the visible puppet that moves the chess pieces] just so high as to bring his eyes above the level of the chess-board. It is very probable that he seats himself upon the little square block or protuberance which is seen in a corner of the main compartment when the doors are open. [100]

Poe then contrasted this pseudo-machine with Babbage's Analytical Engine, whose initial conditions, once specified, allow it to proceed automatically without human intervention, thus making it a real machine:

> Arithmetical or algebraical calculations are, from their very nature, fixed and determinate. Certain *data* being given, certain results necessarily and inevitably follow. These results have dependence upon nothing, and are influenced by nothing but the *data* originally given. And the question to be solved proceeds, or should proceed, to its final determination, by a succession of unerring steps liable to no change, and subject to no modification. This being the case, we can without difficulty conceive the *possibility* of so arranging a piece of mechanism, that upon starting it in accordance with the *data* of the question to be solved, it should continue its movements regularly, progressively, and undeviatingly towards the required solution, since these movements, however complex, are never imagined to be otherwise than finite and determinate. But the case is widely different with the Chess-Player… [100, p. 319, emphasis in the original]

Like Medawar, Brillouin, and Lovelace, Poe is here making the point that information output by a mechanical process cannot exceed or otherwise transcend prior information input.

I have reviewed this history of conservation of information for search, whether called by that name or described in equivalent terms, to show just how deep-seated this idea is. Early accounts of conservation of information applied to deterministic searches based on fully specified data ("axioms" or "premises") and acted upon by fully specified operations ("logical" or "computer-programmable"). Notably lacking from these early accounts were stochastic elements. The outputs of these deterministic searches, if they were assigned probabilities at all, would have probability 0 or 1. The goal of these searches was not to resolve needle-in-a-haystack problems but rather to confirm that certain outputs followed necessarily from certain inputs or else were entirely precluded by those inputs. I want therefore next to fast-forward to the 1990s, when conservation of information was generalized to cover the full range of search, both deterministic and stochastic.

In the 1990s, the biggest result in conservation of information for search was the no-free-lunch theorems of computer scientists David Wolpert and William Macready.

[64] See [98]. Emphasis in the original. Interestingly, Alan Turing spends much precious space in his famous article "Computing Machinery and Intelligence" [99] responding to this quote by Lovelace, arguing that machines such as Babbage's Analytical Engine could nonetheless be said to think provided they reached a certain critical threshold of complexity—presumably as evidenced by passing the imitation game (Turing Test), convincing humans that they are themselves human. But even the claim that machines can think in the same way that humans think (i.e., computational reductionism) does not contradict Lovelace's point, which is that machines in their outputs cannot exceed their inputs. Though I reject that human minds are equivalent to computing machines (i.e., I reject computational reductionism), it seems that a charitable reading of Lovelace credits her with making the same point as Medawar and Brillouin. It is a question of whether machines are limited in their information output by their prior information input. Whether machines can think and what it would mean for them to think are separate questions.

Note that I am using the term "conservation of information for search" here quite broadly, and not in the narrow sense as developed in the body of this paper. Wolpert and Macready's first paper on no free lunch was a Santa Fe Institute technical report titled "No Free Lunch Theorems for Search" [10], positioning these theorems squarely on the side of search (as opposed to physics), and focused on search in general (both deterministic and stochastic). This paper appeared in 1996 as a typescript for limited distribution. Its full formal version appeared subsequently in 1997 in an IEEE journal, where it received a new title: "No Free Lunch Theorems for Optimization" [11]. But optimization is simply a search for something optimal, so the focus of this latter paper was still on search.

The formulation and proof of the no-free-lunch theorems requires some technical firepower from mathematics and computer science, and so a detailed account is not readily comprehensible to non-experts and will not be attempted here. Yet the underlying idea of no free lunch is straightforward and intuitively compelling. As a helpful example that Robert Marks and I have used to illustrate the idea of no free lunch, imagine 52 cards from a standard deck have been arranged at random face down on a table. The challenge is to find the ace of spades A♠. The probability of flipping over A♠ in five or fewer card flips does not depend on any method of flipping the cards. For instance, noting that a previously flipped card is, say, K♣ provides no guidance for the next card to flip over and does not help increase the probability of successfully locating A♠. In five card flips, no matter what method is used to flip them, the probability of find A♠ remains the same—namely, $\frac{5}{52}$ [101].

While this example makes good intuitive sense, it also oversimplifies the no-free-lunch theorems. What we are calling a *method* for finding the target (the target in this example being A♠) is, in the no-free-lunch theorems, a *search algorithm*. Search algorithms in computer science consist of four main components:

1. **an initialization**, defining the starting point of the search and often chosen at random;

2. **a feedback mechanism**, that measures and notes the degree of optimality of each item queried (represented as a fitness function in evolutionary computing and as a parameterization in neural networks[65]);

3. **an update rule**, which uses feedback from the feedback mechanism to determine the next location(s) in the search space to be queried; and

4. **a stop criterion**, specifying when to terminate the search (ideally, the search terminates once it successfully locates the target, but it may also terminate when the search has executed a fixed number of queries or even when it runs out of steam for having exhausted its computational or other scarce resources).

Note that biological evolution, inspired by this conception of search, sees components 2 and 3 as corresponding to differential survival and reproduction.

In terms of the framework for search algorithms as just outlined, the essence of the no-free-lunch theorems is captured in the interplay between components 2 and 3. To see how this works, imagine an algorithm A searching a given search space. To be applicable to A, the no-free-lunch theorems assume that feedback available to A is sufficiently varied and extensive to allow the entire search space to be searched by the algorithm. In other words, any target is searchable anywhere in the search space because, by assumption, feedback is available that enables it to be found.[66] No free lunch then states that as the update rule manages this feedback, search performance for the algorithm A, when averaged across all the different forms that the feedback can take, is constant and equivalent to blind search. What this means is that no search algorithm has an across-the-board advantage over any other search algorithm, and so, in the absence of the right feedback, any search algorithm is no better than blind search—and we need never do

[65] A fitness function in evolutionary computing and a parameterization in neural networks both guide optimization, but in different ways. A fitness function evaluates candidate solutions in evolutionary algorithms, determining their optimality. The evolutionary algorithm then selects the best-performing solutions for the next round of fitness evaluation. In contrast, parameterization in neural networks refers to the weights and biases that define the model's behavior. These parameters are iteratively refined through gradient-based optimization like backpropagation, adjusting continuously based on the gradients of a loss function. While evolutionary computing relies on discrete selection driven by fitness evaluation, neural networks optimize through continuous parameter updates. Both approaches seek improved performance over iterations, but fitness functions operate on whole solutions, whereas parameterizations refine network infrastructure. Evolutionary computing typically explores a discrete or combinatorial solution space, whereas neural networks adjust continuous-valued parameter weights within high-dimensional spaces.

[66] A search algorithm with access to fully varied feedback is thus able to explore the entirety of the search space. This is sometimes expressed by saying that the algorithm is *closed under permutation*. Accordingly, it does not matter how you relabel the points in the search space because any such re-labeling leaves unaffected the effectiveness of a search algorithm in exploring the search space. For details about this point, see [102]. Igel and Toussaint urge that it can be appropriate in biology to restrict the range of feedback, thus making the no-free-lunch theorems inapplicable to certain types of evolutionary searches. Such search restrictions could then improve search success beyond the bounds of no free lunch by finding and exploring salient sectors of biological configuration space. The problem with such restrictions, however, is that they invariably introduce information that in turn needs to be accounted for.

worse than blind search.[67]

Around 2000 I had a conversation with an MIT computer scientist about the predilection of her students at the time to throw an evolutionary algorithm at any problem they might encounter, as though evolutionary algorithms were universal problem solvers having special advantages over other types of algorithms (neural networks nowadays seem to play that role and capture that mystique). She saw this kneejerk application of evolutionary algorithms as naïve, inconsistent with the then recently proven no-free-lunch theorems. To her mind, there was nothing magical about any such algorithms. Rather, their success was unexceptional, dependent not on general properties of the algorithms but rather on effectively adapting the algorithms via their feedback mechanisms to the specific problems at hand. For evolutionary algorithms, fitness functions applied to elements of the search space provided the crucial feedback. And so, the challenge with these algorithms was to find particular fitness functions capable of solving particular problems. Note that in keeping with the Law of Conservation of Information as proven in this paper, to find such a fitness function would be to conduct a successful search for a search.

The choice of algorithm type to use in solving a problem is therefore merely preamble to the real work of adapting and applying the algorithm. These days, with the success of large language models (LLMs), the search algorithm of choice to throw at problems is neural networks. Granted, neural networks are remarkably versatile and boast many successes (natural language processing, computer vision, speech recognition, generative art, game playing, etc.). Yet they are useless unless effectively updated through their feedback mechanisms. No free lunch makes precisely that point: all the interesting work in search algorithms is contained in their feedback mechanisms and, specifically, in the particular feedback that is supplied to the algorithm as it conducts its search. Absent appropriate feedback, no algorithm outperforms blind search. And since blind search is always feasible, blind search sets the baseline for all search.

An equivalent way to understand no free lunch is in terms of biasing. Successful search always requires biasing a search in the direction of a successful outcome, implying that biasing can also occur in the direction of unsuccessful outcomes. And so, when all this biasing is averaged out, no advantage is obtained in finding the object of interest (think of an Easter egg hunt guided by good and bad directions—good directions bias in favor of finding the egg, bad directions bias against finding it). As noted in Section 13, George Montañez has proven a "conservation of bias" result in this vein that extends no free lunch to conservation of information. In any case, because no-free-lunch theorems show that average performance of certain classes of search algorithms is equivalent to performance by blind search, these theorems maintain a strict form of conservation where conservation is not just the best that can be achieved but actually is achieved—albeit at the level of, and never rising above, blind search.

Computer scientist Thomas English, writing about the no-free-lunch theorems of Wolpert and Macready soon after their groundbreaking work first appeared, thus went so far as to redub their results as "conservation of information" results:

> The recent "no free lunch" theorems of Wolpert and Macready indicate the need to reassess empirical methods for evaluation of evolutionary and genetic optimizers. Their main theorem states, loosely, that the average performance of all optimizers is identical if the distribution of functions is average. The present work [is about] the *conservation of information* as an optimizer evaluates points. It is shown that the information an optimizer gains about unobserved values is ultimately due to its prior information of value distributions. Inasmuch as information about one distribution is misinformation about another, there is no generally superior function optimizer. Empirical studies are best regarded as attempts to infer the prior information optimizers have about distributions—i.e., to determine which tools are good for which tasks.[68]

In rounding out this discussion of no free lunch, I should mention a groundbreaking paper on conservation

[67] Blind search explores a search space without domain-specific heuristics or other information to guide the search to its target. Common types of blind search include uniform random sampling (with and without memory) and exhaustive search (traversing the entire search space, or as much of it as is feasible, in no special order). Blind search guarantees finding a solution if enough locations in the search space can be queried, but this is not feasible if the search space is excessively large. Blind search tends to be inefficient in that it is wont to explore large portions of the search space that have no bearing on the target being searched. Unlike heuristic search, which uses problem-specific knowledge/information to direct exploration, blind search treats all possible ways of traversing the search space as equivalent, making it computationally expensive for large-scale problems. While often used as a baseline for comparison, as in the no-free-lunch theorems, blind search works successfully in practice only for search spaces that are not overly large and where little or no information is available to guide search.

[68] From the abstract to [103]. Emphasis added. Note the date, 1996, which means that English was responding to Wolpert and Macready's original Santa Fe Institute paper on no free lunch (1996) and not their subsequent more polished paper for the IEEE (1997). English expanded on the underlying information-theoretic rationale for using the language of conservation of information to characterize no free lunch in [104] and also in [105].

of information (broadly construed) by Cullen Schaffer that appeared in 1994, before all the hubbub surrounding the Wolpert and Macready no-free-lunch theorems two years later. Schaffer's paper focused on conservation for generalization performance [106]. In it, he proves a fundamental result about machine learning, showing that generalization performance—how well a model predicts on unseen data—is inherently a zero-sum game: total generalization performance of a learner, averaged over all learning situations, is zero.

Schaffer showed that for any machine learning model, performance improvements in one subset of learning problems are exactly offset by performance degradation in others. Schaffer's conservation law imposes strict limits on the ability of any learner to perform well universally, emphasizing that gains in generalization are always balanced by losses elsewhere in the problem space. Schaffer explores the implications of this result for the design, evaluation, and enhancement of machine learning models. Even though he shifts the idiom from search to learning, the two can be viewed as equivalent (learning is a search for knowledge). Arguably, all the core elements in Wolpert and Macready's no-free-lunch theorems are prefigured in Schaffer's Conservation Law for Generalization Performance.

With the no-free-lunch theorems, something is clearly being conserved in that performance of different search algorithms, when averaged across a range of feedback mechanisms, is constant and equivalent to performance of blind search. The question then arises how no free lunch relates to the consistent claim in the earlier conservation-of-information literature about output information not exceeding input information. In fact, the connection is straightforward. The only reason to average performance of algorithms across a range of feedback mechanisms is that we do not have any domain-specific information about any particular mechanism to help us, with high probability, find the target in question.

Consequently, no free lunch tells us that without such domain-specific information, we have no special input information to improve the search, and thus no way to achieve output information that exceeds the capacities of blind search. When it comes to search, blind search is always the lowest common denominator—any search under consideration must always be able to perform at least as well as blind search because we can always execute a blind search. With no free lunch, it is blind search as input and blind search as output. Averaging across feedback mechanisms thus acts as a leveler, ensuring parity between information input and output. *No free lunch preserves strict conservation precisely because it sets the bar so low at blind search.*

I have pondered the no-free-lunch theorems since the late 1990s. In that time, I have always found these theorems to be intriguing but also unsatisfying. It is fine and well that search algorithms averaged across a range of different possible feedback mechanisms cannot rise in performance above blind search. The fact is, however, what makes search algorithms interesting is their ability to perform better than blind search. Blind search sets a low bar for search performance that is meant to be surpassed. Wolpert and Macready underscored this very point in the conclusion of their famous 1997 paper, noting that what empowers algorithms to outperform blind search is incorporating the right information, or what they call "problem-specific knowledge" [11].

Classical conservation of information, as developed by Medawar, Brillouin, and others, required that any increase in information output always had to be repaid by information input at least as costly. But classical conservation of information did not put precise numbers on the relation between information input and information output. In fact, it did not have to because its searches depended entirely on logical consequence relations (logical inference rules or logical programming steps) that proceeded deterministically. Its searches tended to be over manageable spaces and could thus afford to be exhaustive. For purely deterministic exhaustive search, the probabilities are all 0 or 1, and success is all or nothing. Classical conservation of information therefore did not require fine gradations for measuring information.

But as soon as searches also incorporated stochastic elements, and thus were no longer deterministic, precise quantitative relations between information input and information output were required for any causally adequate theory of search. A needle-in-a-haystack search with probability p close to 0 for successfully finding the needle requires additional information if that probability is to be raised to a probability q close to 1 sufficient for successfully finding the needle. Conservation of information as developed in the work of Robert Marks, Winston Ewert, George Montañez, and me then established that raising the probability of successful search from p to q incurred a probability cost of p/q, thereby providing a precise numerical accounting of how information output in search is related to prior information input.[69]

This earlier work on conservation of information showed that attempts to enhance search success (such as by turning a near-impossible search into a feasible one) merely shifted the informational burden but did not eliminate it. This earlier work was adapted on a case-by-case basis to specific types of search and therefore lacked full generality, as is evident from the catalogue of earlier conservation of information theorems recounted in Appendix 2. The present paper establishes a unifying probabilistic framework that generalizes all the previous

[69] For an overview of this work, see the publications page of the Evolutionary Informatics Lab: https://www.evoinfo.org/publications.html (last accessed January 16, 2025).

work on conservation of information. By distilling its findings into a single fundamental relation of probability theory, this work provides a definitive, fully developed, general formulation of the Law of Conservation of Information, showing that information that facilitates search cannot magically materialize out of nothing but instead must be derived from pre-existing sources.

APPENDIX 2: EARLIER CONSERVATION-OF-INFORMATION THEOREMS

The Law of Conservation of Information as proven in this paper is the most general conservation-of-information theorem that exists. In fact, it is as general as it could be given the fundamental probabilistic claim that conservation of information makes about the relation between information inputs and information outputs in search. Yet ironically, even though it is more general and powerful than earlier conservation-of-information theorems, it is also simpler to prove and involves less advanced mathematical methods than the earlier theorems. In fact, if one presupposes its Bayesian underpinnings to justify and coordinate its probabilities, the conservation-of-information theorem proven in this paper follows, as saw in Sections 9 and 10, directly from the very basic probabilistic inequality $P(E) \geq P(E|F) \times P(F)$.

Even though the earlier conservation-of-information theorems are now special cases of a more general and powerful result, it is worth reviewing them for historical completeness. Additionally, examining these theorems helps to clarify how the Law of Conservation of Information, as proven in this paper, does indeed simplify and extend our understanding of conservation of information. Insofar as this paper attempts to provide a comprehensive theoretical treatment of conservation of information, the historical context provided by these earlier theorems deserves to be part of that treatment. Moreover, even though these earlier theorems are now superseded, they have and continue to generate a growing literature in the intelligent design community, as indicated in Section 13.

In what follows next, I will simply state four conservation-of-information theorems. I will first provide some minimal commentary before stating the theorems by way of a preamble. But I will not provide proofs. The proofs can be found in the endnote references given with the statement of each theorem. Also, the formulation of these theorems uses information measures rather than probability measures. As noted at the end of Section 5, information measures are probability measures transformed by taking a negative logarithm to the base 2, and so baseline probabilities become endogenous information, improved probabilities become exogenous information, and S4S probabilities become active information.

The first three theorems below are all formulated in terms of uniform probabilities. The first two—namely, the *function-theoretic* and *fitness-theoretic* versions—are the simplest in terms of mathematical sophistication, requiring for their proofs only elementary combinatorics. The third, the *basic measure-theoretic* version, requires measure theory and functional analysis to prove, but it is likewise focused on uniform probabilities (both for the original search space and for the S4S space). The fourth, the *general measure-theoretic* version, requires all the mathematics of the basic version and makes in its proof extensive use of vector-valued integration.

Of these various versions, the first two are mainly of academic interest, helping to provide an intuitive understanding of conservation of information but without widespread applications. The function-theoretic version illustrates that narrowing down on a target by excluding possibilities outside the target is an informational act that adds information and obeys conservation of information. The fitness-theoretic version shows that subcollections of fitness functions that increase the probability of finding a target constitute higher-level targets within the full collection of fitness functions (the S4S space), and that the S4S probability or active information of such higher-level targets obeys conservation of information. The fitness-theoretic version is of interest also because it incorporates a very natural no-free-lunch theorem in which the performance metric is cashed out simply in terms of the probability of the desired outcome, and where averaging over the fitness functions leads to performance of a blind search as characterized by uniform probabilities.

The basic and general measure-theoretic versions identify searches with the probabilities (or information values) that searches confer on finding targets. Thus the search-for-a-search space itself becomes a collection of probability measures. This is convenient and useful for applying these theorems to explain increases in information that arise in scientific investigations, especially in biological evolution and cosmological fine-tuning. Yet because the basic measure-theoretic version is limited to working solely with uniform probabilities on the search and search-for-a-search spaces, its applicability to such problems is limited (though not completely undercut because uniform probabilities come up widely and often provide a helpful first approximation to the more exact underlying probabilities). In any case, the general measure-theoretic version avoids the limitations of the basic version by also covering probability measures that arise in search quite generally—uniform and non-uniform.

As a conservation-of-information theorem, the general measure-theoretic version requires quite a bit of mathematical machinery to prove, certainly more than the Law of Conservation of Information as proven in this paper. Given that the latter is at once less math intensive and conceptually more powerful than the former, it may be thought that the general measure-theoretic version is superfluous, and that nothing would have been lost by simply proving, right out of the blocks, the Law of Conservation of Information as articulated in this paper. But in fact, the general measure-theoretic version, when applicable, can helpfully delineate the search-for-a-search space. Depending on the search context, the general measure-theoretic version can fill in the underlying nuts and bolts of a search and its related search for a search that our Bayesian framework for the Law of Conservation of Information can conveniently ignore.

Note that all the following theorems are formulated for finite search spaces. Their proofs all get the ball rolling with combinatorial calculations over the initial search space. The function-theoretic and fitness-theoretic versions then carry this finiteness over to the corresponding search-for-a-search space of each. For the measure-theoretic versions, on the other hand, the corresponding search-for-a search space of each ends up being a space of probability measures that is isomorphic to a finite dimensional simplex in Euclidean space. Consequently, this higher-order search space is, for these versions, infinite. In all of the following theorems, finiteness of the underlying search space never proved a problem in applications because infinite probability spaces can for all practical purposes always be approximated to any degree of accuracy by finite probability spaces.[70] Nonetheless, the Law of Conservation of Information as formulated in this paper has the advantage that it applies regardless of whether the search spaces are finite or infinite —the theory developed here covers both cases.

Before stating these four conservation-of-information theorems, it bears repeating that other theorems exist under the name "conservation of information." Yet the focus of these other theorems differs from those listed below. In Appendix 1, I examined the history of conservation of information, noting that Thomas English in effect redubbed the no-free-lunch theorems as conservation-of-information theorems. There I also cited Cullen Schaffer's work on conservation for generalization performance as a proto-no-free-lunch result. In Section 13, I reviewed conservation-of-information results in the work of George Montañez. While Montañez's results expand the notion of conservation of information, they do not, like the results below, track information costs for fixed targets. None of this is meant to question the significance of the work of English, Schaffer, Montañez, or others on conservation of information broadly construed. But for the Law of Conservation of Information as proven in this paper, its lineal descent is clear, which is to say the four following theorems, and in particular the last two. Note that in what follows, the logarithmic function log is the logarithm to the base 2 (namely, $\log_2$).

[70] See Billingsley, Patrick, *Convergence of Probability Measures*, 1st ed. (New York: Wiley, 1968). Note that a second edition appeared in 1999, but it is riddled with typos and introduces a revamped notation that is confusing.

Conservation of Information Theorem (function-theoretic version) [107, pp. 374–376]. Let T be a target in Ω. Assume that Ω is finite and nonempty, and (using enclosing vertical bars to denote the number of items in a set) that $p = |T|/|\Omega|$ (which we take to be extremely small). The endogenous information is therefore $I_\Omega = -\log(p)$. Next, let Ω' be another nonempty finite space, φ be a function that maps Ω' to Ω, and $T' = \{y \in \Omega' \mid \varphi(y) \in T\}$. Or, in standard set-theoretic notation, $T' = \varphi^{-1}(T)$. Define $q = |T'|/|\Omega'|$ (which we take to be considerably bigger than p). Given a null [baseline] search for T' in Ω', φ induces an alternative search S for T in Ω. The exogenous information is therefore $I_S = -\log(q)$. Next, define $\mathcal{F}$ as the set of all functions from Ω' to Ω and $\mathcal{T}$ as the set of all functions ψ in $\mathcal{F}$ such that $\left|\psi^{-1}(T)\right|/|\Omega'| \geq q$ (i.e., each such ψ maps at least as many elements of Ω' to T as φ). Then $|\mathcal{T}|/|\mathcal{F}| \leq p/q$, or equivalently the (higher-order) endogenous information associated with finding $\mathcal{T}$ in $\mathcal{F}$, i.e., $-\log(|T|/|F|)$, is bounded below by the (lower-order) active information $I_+ = \log(q/p)$. [Note that the S4S space here is the function space $\mathcal{F}$.]

Conservation of Information Theorem (fitness-theoretic version) [107, pp. 378–381]. Let T be a target in Ω. Assume Ω is finite and T is nonempty. Let $\mathbf{U}$ denote the uniform probability distribution on Ω and let $p = |T|/|\Omega| = \mathbf{U}(T)$(which we take to be extremely small). The endogenous information is therefore $I_\Omega = -\log(p)$. Next, let $\mathcal{F}$ denote a finite collection of fitness functions on Ω and let $\mathbf{S_\Omega}$ denote the symmetric group on Ω (i.e., all permutations of this set). Without loss of generality, assume that any f in $\mathcal{F}$ only takes values in $\{0, 1/M, 2/M, ..., (M–1)/M, 1\}$ for some large fixed M and that $\mathcal{F}$ includes all such f. $\mathcal{F}$ is therefore closed under the symmetric group $\mathbf{S_\Omega}$, i.e., for any f in $\mathcal{F}$ and any σ in $\mathbf{S}_\Omega$, $f \circ \sigma$ is also in $\mathcal{F}$. Suppose further that any f in $\mathcal{F}$ induces a probability distribution P_f on Ω (corresponding to an alternative search). Assume that each such P_f satisfies the following invariance property: for any σ in $\mathbf{S}_\Omega$ and $A \subset \Omega$, $P_{f \circ \sigma}(\sigma^{-1}A) = P_f(A)$. An NFL result then follows:

$$\frac{1}{|\mathcal{F}|}\sum_{f \in \mathcal{F}} P_f(T) = \mathbf{U}(T).$$

Assume next that g in $\mathcal{F}$ is such that $q = P_g(T)$ (which we take to be considerably bigger than p) and that g induces an alternative search S for which the exogenous information is $I_S = -\log(q)$. Let $\mathcal{T}$ denote the set of all fitness functions h in $\mathcal{F}$ such that $P_h(T) \geq q$ (i.e., each such h induces a probability distribution that assigns at least as much probability to T as P_g; equivalently, each such h induces an alternative, or evolutionary, search at least as effective at locating T as S). Then, the (higher-order) uniform probability of $\mathcal{T}$ in $\mathcal{F}$, i.e., $|\mathcal{T}|/|\mathcal{F}|$, which may be denoted by $\mathbf{U}^*(\mathcal{T})$, is less than or equal to p/q. Equivalently, the (higher-order) endogenous information associated with finding $\mathcal{T}$ in $\mathcal{F}$, i.e., $-\log(\mathbf{U}^*(\mathcal{T}))$, is bounded below by the (lower-order) active information $I_+ = -\log(\mathbf{U}(T)) + \log(P_g(T)) = \log(q/p)$. [Note that the S4S space here is the space of fitness functions $\mathcal{F}$.]

Conservation of Information Theorem (basic measure-theoretic version)[71] Let T be a target in Ω. Assume Ω is finite and T is nonempty. Let $\mathbf{U}$ denote the uniform probability distribution on Ω and let $p = \frac{|T|}{|\Omega|} = \mathbf{U}(T)$ (which we take to be extremely small). The endogenous information is therefore $I_\Omega = -\log(p)$. Next, let μ be a probability distribution on Ω such that $q = \mu(T)$ (which we take to be considerably bigger than p). Suppose that μ characterizes the probabilistic behavior of an alternative search S. The exogenous information is therefore $I_S = -\log(q)$. Next, let $\mathcal{M}$ denote the set of all probability distributions on Ω and $\mathcal{T}$ be the set of probability distributions ν in $\mathcal{M}$ such that $\nu(T) \geq q$ (i.e., each such ν assigns at least as much probability to T as μ—each such ν therefore represents a search that is at least as effective at locating T as μ). Then the (higher-order) uniform probability of $\mathcal{T}$ in $\mathcal{M}$, which may be denoted by $\mathbf{U}^*(T)$, is less than or equal to p/q. Equivalently, the (higher-order) endogenous information associated with finding $\mathcal{T}$ in $\mathcal{M}$, i.e., $-\log(\mathbf{U}^*(\mathcal{T}))$, is bounded below by the (lower-order) active information $I_+ = -\log(\mathbf{U}(T)) + \log(\mu(T)) = \log(\frac{q}{p})$.

Conservation of Information Theorem (general measure-theoretic version).[72] Let T be a target in Ω. Assume Ω is finite and T is nonempty. Let $\mathbf{U}$ denote the uniform probability distribution on Ω and let $p = \frac{|T|}{|\Omega|} = \mathbf{U}(T)$ (which we take to be extremely small). Next, let μ and ν be probability measures on Ω such that $q = \mu(T)$ and $r = \nu(T)$. We assume that $p \leq q < r$. Suppose that μ characterizes the probabilistic behavior of a search S and that ν characterizes the probabilistic behavior of a search S'. We treat μ as the null search and ν as the alternative search, thus making μ the natural probability associated with Ω. Accordingly, $I_S = -\log(q)$ becomes the endogenous information and $I_{S'} = -\log(r)$ the exogenous information. [Note that contrary to the notation in this paper, q is now a small probability, close to p, and r is a much larger probability, close to 1.] On the space of all probability measures on Ω—namely, $\mathbf{M}(\Omega)$—there is then a higher-order target induced by T with respect to r—namely, $\overline{T}_r = \{\theta \in \mathbf{M}(\Omega) | \theta(T) \geq r\}$. The higher-order probability measure induced by μ on $\mathbf{M}(\Omega)$—namely, $\bar{\mu}$—then yields the probability $\bar{\mu}(\overline{T}_r)$, which is less than or equal to q/r. Equivalently, the (higher-order) endogenous information associated with finding the (higher-order) target $\overline{T}_r$ in $\mathbf{M}(\Omega)$, i.e., $-\log(\bar{\mu}(\overline{T}_r))$, is bounded below by the (lower-order) active information $I_+ = -\log(\mu(T)) + \log(\nu(T)) = \log(\frac{r}{q})$.

[71] For the statement of this result and an outline of the proof, see [107, pp. 376–378]. For the full proof see Appendix C of [48].

[72] See [49, pp. 59–60]. Note that the proof of the general case depends on the proof of the basic case for uniform probabilities in that the probabilities in the general case are probability densities integrated with respect to uniform probabilities—uniform probabilities that exist both on the original search space and on the search-for-a-search space.

ACKNOWLEDGMENTS

I first started exploring the idea of conservation of information in the late 1990s as I undertook to analyze Richard Dawkins' METHINKS IT IS LIKE A WEASEL computer simulation. At the time, Paul Nelson alerted me to the then recent work by David Wolpert and Bill Macready on no-free-lunch theorems. Building on this work, I developed the notions of displacement and conservation of information. Yet my early treatment of these concepts was, as philosophers say, "pre-theoretic." The right intuitions were there, but not a precise formal treatment.

Between 2005 and 2010, in collaboration with Robert Marks at Baylor, and then with his graduate students Winston Ewert and George Montañez, we were able to push these ideas from the pre-theoretic to the theoretic stage. And yet conservation of information in a formal mathematical sense only applied to specific cases (particular probabilistic models and distributions).

For me, the eureka moment regarding conservation of information occurred at the Evangelical Philosophical Society meeting in San Diego, November 21, 2024. In a symposium at that meeting organized by Melissa Cain-Travis and moderated by John Bloom, Douglas Axe gave a talk in which he illustrated how raising the probability of a successful search could be represented in terms of Bayesian conditional probabilities. That is when it became clear to me that there is indeed a Law of Conservation of Information that holds with perfect generality.

Grappling with conservation of information has been for me a long journey. With this paper, the Law of Conservation of Information is now rigorously justified, having a strong theoretical foundation and a broad sphere of application. I want to thank all the people mentioned above as well as colleagues and friends at Discovery Institute who have supported my work in this area over the years. Here I would like especially to express my gratitude to Steve Meyer, John West, Casey Luskin, Bruce Gordon, Mike Behe, Brian Miller, and the late Jonathan Wells.

Finally, I would like to thank five reviewers for comments that helped greatly improve this paper.

1. Lopes Coelho R (2009) On the Concept of Energy: How Understanding its History can Improve Physics Teaching. Science & Education 18:961–983. doi:10.1007/s11191-007-9128-0
2. Marks II RJ, Dembski WA, Ewert W (2017) Introduction to Evolutionary Informatics. World Scientific.
3. Gleick J (2011) The Information: A History, a Theory, a Flood. Pantheon Books, New York, 1st edition.
4. Shannon CE, Weaver W (1949) The Mathematical Theory of Communication. University of Illinois Press, Urbana.
5. Cover TM, Thomas JA (2006) Elements Of Information Theory. Wiley-Interscience, Hoboken, NJ.
6. Stone LD, editor (2010) Theory of Optimal Search. Number v. 118 in Mathematics in Science and Engineering. Academic Press, New York.
7. Chudnovsky D, Chudnovsky G, editors (1989) Search Theory: Some Recent Developments. Number 112 in Lecture Notes in Pure and Applied Mathematics. Marcel Dekker, New York.
8. Aydinian H, Cicalese F, Deppe C (2013) Information Theory, Combinatorics, and Search Theory: In Memory of Rudolf Ahlswede. Number 7777 in Lecture Notes in Computer Science. Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-36899-8
9. Stone LD, Royset JO, Washburn AR (2016) Optimal Search for Moving Targets, volume 237 of International Series in Operations Research & Management Science. Springer International Publishing, Cham. doi:10.1007/978-3-319-26899-6
10. Wolpert DH, Macready WG (1995) No Free Lunch Theorems for Search. Working paper, Santa Fe Institute.
11. Wolpert DH, Macready WG (1997) No Free Lunch Theorems for Optimization. IEEE Transactions on Evolutionary Computation 1:67–82. doi:10.1109/4235.585893
12. Culberson JC (1998) On the Futility of Blind Search: An Algorithmic View of "No Free Lunch". Evolutionary Computation 6:109–127. doi:10.1162/evco.1998.6.2.109
13. Dembski WA, Marks II RJ (2009) Conservation of Information in Search: Measuring the Cost of Success. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 39:1051–1061. doi:10.1109/TSMCA.2009.2025027
14. Polkinghorne JC (2005) Exploring Reality: The Intertwining of Science and Religion. Yale University Press, New Haven.
15. Dennett DC (1996) Darwin's Dangerous Idea: Evolution and the Meanings of Life. A Touchstone Book. Simon & Schuster, New York, 1. touchstone edition.
16. Kauffman SA (1996) At Home in the Universe: The Search for Laws of Self-Organization and Complexity. Oxford university press, New York [etc.].
17. Kauffman SA (2000) Investigations. Oxford Univ. Press, Oxford.
18. Dembski WA (2002) No Free Lunch : Why Specified Complexity Cannot Be Purchased without Intelligence. Rowman & Littlefield, Lanham MD.
19. Dembski WA, Ewert W (2023) The Design Inference: Eliminating Chance through Small Probabilities. Discovery Institute, Seattle, 2nd edition.
20. Behe MJ (2020) A Mousetrap for Darwin. Discovery Institute Press, Seattle, WA, first edition.
21. Dawkins R (1986) The Blind Watchmaker: Why The Evidence Of Evolution Reveals A Universe Without Design. Norton, New York.
22. Jacobson SH, Yücesan E (2004) Analyzing the Performance of Generalized Hill Climbing Algorithms. Journal of Heuristics 10:387–405. doi:10.1023/B:HEUR.0000034712.48917.a9
23. Dawkins R (1997) Climbing Mount Improbable. W.W. Norton & Company, New York.
24. Lenski RE, Ofria C, Pennock RT, Adami C (2003) The evolutionary origin of complex features. Nature 423:139–44. doi:10.1038/nature01568
25. Pennock RT (2004) DNA by Design?: Stephen Meyer and the Return of the God Hypothesis. In WA Dembski, M Ruse, editors, Debating Design, 130–148. Cambridge University Press, 1 edition. doi:10.1017/CBO9780511804823.008
26. Schneider TD (2000) Evolution of Biological Information. Nucleic Acids Research 28:2794–2799. doi:10.1093/nar/28.14.2794
27. Thomas D (2010) War of the Weasels: An Evolutionary Algorithm Beats Intelligent Design. Skeptical Inquirer 43:42–46.
28. Aigner M (2007) Discrete Mathematics. American Mathematical Society, Providence, 1st edition.
29. Schervish MJ (1997) Theory of Statistics. Springer Series in Statistics. Springer, New York, corr. 2nd print edition.
30. Bogachev VI (2018) Weak Convergence of Measures. Number volume 234 in Mathematical Surveys and Monographs. American Mathematical Society, Providence, Rhode Island.
31. Polanyi M (1967) Life Transcending Physics and Chemistry. Chemical & Engineering News Archive 45:54–69. doi:10.1021/cen-v045n035.p054
32. Lenoir T (1989) The Strategy of Life: Teleology and Mechanics in Nineteenth-Century German Biology. University of Chicago Press, Chicago.
33. Yockey HP (1992) Information Theory and Molecular Biology. Cambridge University Press, Cambridge ; New York, NT, USA.
34. Meyer SC (2003) DNA and the Origin of Life: Information, Specification, and Explanation. In J Campbell, S Meyer, editors, Darwinism, Design and Public Education. Michigan State University Press, Lansing, MI.
35. Paley W (1802) Natural Theology; or, Evidences of the Existence and Attributes of the Deity, Collected from the Appearances of Nature. R. Faulder, London.
36. Dembski WA (2004) The Design Revolution: Answering The Toughest Questions About Intelligent Design. InterVarsity Press.
37. Gibbons A (1985) Algorithmic Graph Theory. Cambridge University Press, Cambridge [Cambridgeshire] ; New York.
38. Lucretius (1924) De Rerum Natura, volume 5 of Loeb Classical Library. G. P. Putnam's Sons.
39. Ridley M (2004) Evolution. Blackwell Pub, Malden, MA, 3rd edition.
40. Axe DD (2004) Estimating the prevalence of protein sequences adopting functional enzyme folds. Journal of molecular biology 341:1295–315. doi:10.1016/j.jmb.2004.06.058
41. Miller BJ (2024) A percolation theory analysis of continuous functional paths in protein sequence space affirms previous insights on the optimization of proteins for adaptability. PLOS ONE 19:e0314929. doi:10.1371/journal.pone.0314929
42. Mittelbach AA, Fischlin M (2021) The Theory of Hash Functions and Random Oracles: An Approach to Modern Cryptography. Information Security and Cryptography. Springer, Cham, Switzerland.
43. Gates B, Myhrvold N, Rinearson P (1996) The Road Ahead. Penguin Books, New York, completely rev. and up-to-date edition.
44. Ruse M (1996) Monad to Man: The Concept of Progress in Evolutionary Biology. Harvard University Press, Cambridge, Mass.

45. Penrose R (1999) The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics. Oxford Univ. Press, Oxford.
46. Behe MJ (2007) The Edge of Evolution. Free Press, New York, New York, USA.
47. Darwin C (1859) On the Origin of the Species by Means of Natural Selection: Or, The Preservation of Favoured Races in the Struggle for Life.
48. Dembski WA, Marks II RJ (2010) The Search for a Search: Measuring the Information Cost of Higher Level Search. Journal of Advanced Computational Intelligence and Intelligent Informatics 14:475–486.
49. Dembski WA, Ewert W, Marks II RJ (2013) A General Theory of Information Cost Incurred by Successful Search. Biological Information 26–63. doi:10.1142/9789814508728_0002
50. Díaz-Pachón DA, Hössjer O, Marks Ii RJ (2021) Is cosmological tuning fine or coarse? Journal of Cosmology and Astroparticle Physics 2021:020. doi:10.1088/1475-7516/2021/07/020
51. Díaz-Pachón DA, Hössjer O (2022) Assessing, Testing and Estimating the Amount of Fine-Tuning by Means of Active Information. Entropy 24:1323. doi:10.3390/e24101323
52. Díaz-Pachón DA, Hössjer O, Mathew C (2024) Is It Possible to Know Cosmological Fine-tuning? The Astrophysical Journal Supplement Series 271:56. doi:10.3847/1538-4365/ad2c88
53. Díaz Pachón DA, Marks R (2020) Active Information Requirements for Fixation on the Wright-Fisher Model of Population Genetics. BIO-Complexity 2020. doi:10.5048/BIO-C.2020.4
54. Díaz–Pachón DA, Sáenz JP, Rao JS (2020) Hypothesis testing with active information. Statistics & Probability Letters 161:108742. doi:10.1016/j.spl.2020.108742
55. Thorvaldsen S, Hössjer O (2020) Using statistical methods to model the fine-tuning of molecular machines and systems. Journal of Theoretical Biology 501:110352. doi:10.1016/j.jtbi.2020.110352
56. Thorvaldsen S, Hössjer O (2024) Use of directed quasi-metric distances for quantifying the information of gene families. BioSystems 243:105256. doi:10.1016/j.biosystems.2024.105256
57. Thorvaldsen S, Øhrstrøm P, Hössjer O (2024) The Representation, Quantification, and Nature of Genetic Information. Synthese 204:1–38. doi:10.1007/s11229-024-04613-z
58. Montañez GD (2017) The famine of forte: Few search problems greatly favor your algorithm. In 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 477–482. doi:10.1109/SMC.2017.8122651
59. Montanez GD (2018) A Unified Model of Complex Specified Information. BIO-Complexity 2018. doi:10.5048/bio-c.2018.4
60. Montañez GD, Hayase J, Lauw J, Macias D, Trikha A, Vendemiatti J (2019) The Futility of Bias-Free Learning and Search. In J Liu, J Bailey, editors, AI 2019: Advances in Artificial Intelligence, 277–288. Springer International Publishing, Cham. doi:10.1007/978-3-030-35288-2_23
61. Montañez GD, Bashir D, Lauw J (2021) Trading Bias for Expressivity in Artificial Learning. In AP Rocha, L Steels, J Van Den Herik, editors, Agents and Artificial Intelligence, volume 12613, 332–353. Springer International Publishing, Cham. doi:10.1007/978-3-030-71158-0_16
62. Moore D, Walker SI, Levin M (2017) Cancer as a disorder of patterning information: Computational and biophysical perspectives on the cancer problem. Convergent Science Physical Oncology 3:043001. doi:10.1088/2057-1739/aa8548
63. McMillen P, Walker SI, Levin M (2022) Information Theory as an Experimental Tool for Integrating Disparate Biophysical Signaling Modules. International Journal of Molecular Sciences 23:9580. doi:10.3390/ijms23179580
64. Dingjan T, Futerman AH (2021) The fine-tuning of cell membrane lipid bilayers accentuates their compositional complexity. BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology 43:e2100021. doi:10.1002/bies.202100021
65. Dingjan T, Futerman AH (2024) Fine-tuned protein-lipid interactions in biological membranes: Exploration and implications of the ORMDL-ceramide negative feedback loop in the endoplasmic reticulum. Frontiers in Cell and Developmental Biology 12. doi:10.3389/fcell.2024.1457209
66. Floridi L (2011) The Philosophy of Information. Oxford University Press, Oxford ; New York.
67. Floridi L (2025) AI as Agency without Intelligence: On Artificial Intelligence as a New Form of Artificial Agency and the Multiple Realisability of Agency Thesis. Philosophy & Technology 38:30. doi:10.1007/s13347-025-00858-9
68. Floridi L (2023) AI as Agency Without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models. Philosophy & Technology 36:15. doi:10.1007/s13347-023-00621-y
69. Meyer SC (2023) Return of the God Hypothesis: Three Scientific Discoveries That Reveal the Mind behind the Universe. HarperOne, an imprint of HarperCollinsPublishers, New York, NY, first harpercollins paperback edition.
70. Axe D (2016) Undeniable: How Biology Confirms Our Intuition That Life Is Designed. HarperCollins Publishers, New York, 1st edition.
71. Kamath U, Keenan K, Somers G, Sorenson S (2024) Large Language Models: A Deep Dive: Bridging Theory and Practice. Springer, Cham.
72. Coiffier J (2011) Fundamentals of Numerical Weather Prediction. Cambridge University Press, Cambridge New York.
73. Gen M, Cheng R (1999) Genetic Algorithms and Engineering Optimization. Wiley, 1 edition. doi:10.1002/9780470172261
74. Kamath U, Liu Z, Whitaker J (2019) Deep Learning for NLP and Speech Recognition. Springer, Cham.
75. Craske MG (2017) Cognitive-Behavioral Therapy. Theories of Psychotherapy Series. American Psychological Association, Washington, DC, second edition.
76. Zuckerman G (2019) The Man Who Solved the Market. Portfolio / Penguin, New York, NY.
77. Gerhardt MJ (2008) The Power of Precedent. Oxford University Press, Oxford ; New York.
78. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T (2018) The rise of deep learning in drug discovery. Drug Discovery Today 23:1241–1250. doi:10.1016/j.drudis.2018.01.039
79. Wigner EP (1960) The Unreasonable Effectiveness of Mathematics in the Natural Sciences. Communications in Pure and Applied Mathematics 13:1–14.
80. Rolston H (1999) Genes, Genesis, and God: Values and Their Origins in Natural and Human History. Cambridge University Press, Cambridge, U.K. ; New York.
81. Lewis DK (2001) On the Plurality of Worlds. Blackwell Publishers, Malden, Mass.
82. Gonzalez G, Richards JW (2024) The Privileged Planet: How Our Place in the Cosmos Is Designed for Discovery. Regnery Pub. ; National Book Network [distributor], Washington, DC : Lanham, MD, 20th anniversay edition.
83. Robertson DS (1999) Algorithmic information theory, free will, and the Turing test. Complexity 4:25–34. doi:10.1002/(SICI)1099-0526(199901/02)4:3¡25::AID-CPLX5¿3.0.CO;2-E
84. Moyers B (1991) Introduction. Anchor Books Doubleday, New York.
85. McGilchrist I (2019) The Master and His Emissary: The Divided Brain and the Making of the Western World. Yale University Press, New Haven London, new expanded edition.
86. McGilchrist I (2021) The Matter With Things: Our Brains, Our Delusions, and the Unmaking of the World. Perspectiva Press, London.

87. Bostrom N (2014) Superintelligence: Paths, Dangers, Strategies. Oxford University Press, Oxford, first edition.
88. Leibniz GW (2014) Leibniz's Monadology: A New Translation and Guide. Edinburgh University Press, Edinburgh.
89. Egnor M, O'Leary D (2025) The Immortal Mind. Worthy, New York.
90. Oden TC (2009) Classic Christianity: A Systematic Theology. HarperOne, New York, NY.
91. Susskind L, Lindesay J (2005) An Introduction to Black Holes, Information and the String Theory Revolution: The Holographic Universe. World Scientific, Hackensack, NJ.
92. Kornfel'd IP, Fomin SV, Sinaĭ IG (1982) Ergodic theory. Number 245 in Grundlehren der mathematischen Wissenschaften. Springer-Verlag, New York.
93. Braunstein SL, Pati AK (2007) Quantum Information Cannot Be Completely Hidden in Correlations: Implications for the Black-Hole Information Paradox. Physical Review Letters 98:080502. doi:10.1103/PhysRevLett.98.080502
94. Suppes P (1984) Probabilistic Metaphysics. Blackwell, Oxford, UK ; New York, NY, USA.
95. Medawar PB (1985) The Limits of Science. Oxford Univ. Press, Oxford.
96. Rowland HA (1899) The Highest Aim of the Physicist. Science 10:825–833. doi:10.1126/science.10.258.825
97. Brillouin L (2004) Science and Information Theory. Dover Phoenix Editions. Dover Publications, Mineola, N.Y, 2nd edition.
98. Lovelace A (1843) Translator's notes on menebrea's memoir. In R Taylor, editor, Scientific Memoirs, Selected from the Transactions of Foreign Academies of Science and Learned Societies, and from Foreign Journals, volume 3, 691–731. Richard and John E. Taylor, London.
99. Turing AM (1950) Computing Machinery and Intelligence. Mind A Quarterly Review of Psycology and Philosophy LIX.
100. Poe EA (1836) Maelzel's chess-player. Southern Literary Messenger 2:318–326.
101. Dembski WA, Marks II RJ (2009) Bernoulli's Principle of Insufficient Reason and Conservation of Information in Computer Search. IEEE. doi:10.1109/ICSMC.2009.5346119
102. Igel C, Toussaint M (2005) A No-Free-Lunch theorem for non-uniform distributions of target functions. Journal of Mathematical Modelling and Algorithms 3:313–322. doi:10.1007/s10852-005-2586-y
103. English TM (1996) Evaluation of Evolutionary and Genetic Optimizers: No Free Lunch. In Evolutionary Programming, 163–169.
104. English T (1999) Some information theoretic results on evolutionary optimization. In Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406), volume 1, 788–795 Vol. 1. doi:10.1109/CEC.1999.782013
105. English T No more lunch: Analysis of sequential search. Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No.04TH8753) 227–234. doi:10.1109/CEC.2004.1330861
106. Schaffer C (1994) A Conservation Law for Generalization Performance. In W Cohen, H Willian, editors, Proceedings of the Eleventh International Machine Learning Conference, 259–265.
107. Dembski WA, Marks II RJ (2009) Life's Conservation Law: Why Darwinian Evolution Cannot Create Biological Information. In B Gordon, WA Dembski, editors, The Nature of Nature. ISI Books, Wilmington, Del.