

Probability, Statistics, Evolution, and Intelligent Design

Peter Olofsson

In the last decades, arguments against Darwinian evolution have become increasingly sophisticated, replacing Creationism by Intelligent Design (ID) and the book of Genesis by biochemistry and mathematics. As arguments claiming to be based in probability and statistics are being used to justify the anti-evolution stance, it may be of interest to readers of *CHANCE* to investigate methods and claims of ID theorists.

Probability, Statistics, and Evolution

The theory of evolution states in part that traits of organisms are passed on to successive generations through genetic material and that modifications in genetic material cause changes in appearance, ability, function, and survival of organisms. Genetic changes that are advantageous to successful reproduction over time dominate and new species evolve. Charles Darwin (1809–1892) is famously credited with originating and popularizing the idea of speciation through gradual change after observing animals on the Galapagos Islands.

Today, the theory of evolution is the scientific consensus concerning the development of species, but is nevertheless routinely challenged by its detractors. The National Academy of Sciences and Institute of Medicine (NAS/IM) recently issued a revised and updated document, titled "Science, Evolution, and Creationism," that describes the theory of evolution and investigates the relation between science and religion. Although the latter topic is of interest in its own right, in fairness to ID proponents, it should be pointed out that many of them do not employ religious arguments against evolution and this article does not deal with issues of faith and religion.

How do probability and statistics enter the scene?

In statistics, hypotheses are evaluated with data collected in a way that introduces as little bias as possible and with as much precision as possible. A hypothesis suggests what we would expect to observe or measure, if the hypothesis were true. If such predictions do not agree with the observed data, the hypothesis is rejected and more plausible hypotheses are suggested and evaluated. There are many statistical techniques and methods that may be used, and they are all firmly rooted in the theory of probability, the "mathematics of chance."

An ID Hypothesis Testing Challenge to Evolution

In his book *The Design Inference*, William Dembski introduces the "explanatory filter" as a device to rule out chance explanations and infer design of observed phenomena. The filter also appears in his book *No Free Lunch*, where the description differs slightly. In essence, the filter is a variation on statistical hypothesis testing with the main difference being that it aims at ruling out chance altogether, rather than just a specified null hypothesis. Once all chance explanations have been ruled out, 'design' is inferred. Thus, in this context, design is merely viewed as the complement of chance.

To illustrate the filter, Dembski uses the example of Nicholas Caputo, a New Jersey Democrat who was in charge of putting together the ballots in his county. Names were to be listed in random order, and, supposedly, there is an advantage in having the top line of the ballot. As Caputo managed to place a Democrat on the top line in 40 out of 41 elections, he was suspected of cheating. In Dembski's terminology, cheating now plays the role of design, which is inferred by ruling out chance.

Let us first look at how a statistician might approach the Caputo case. The way in which Caputo was supposed to draw names gives rise to a null hypothesis $H_0 : p = 1/2$ and an alternative hypothesis $H_A : p > 1/2$, where p is the probability of drawing a Democrat. A standard binomial test of $p = 1/2$ based on the observed relative frequency $\hat{p} = 40/41 \approx 0.98$ gives a solid rejection of H_0 in favor of H_A with a p -value of less than 1 in 50 billion, assuming independent drawings. A statistician could also consider the possibility of different values of p in different drawings, or dependence between listings for different races.

What then would a 'design theorist' do differently? To apply Dembski's filter and infer design, we need to rule out all chance explanations; that is, we need to rule out both H_0 and H_A . There is no way to do so with certainty, and, to continue, we need to use methods other than probability calculations. Dembski's solution is to take Caputo's word that he did not use a flawed randomization device and conclude that the only relevant chance hypothesis is H_0 . It might sound questionable to trust a man who is charged with cheating, but as it hardly makes a difference to the case whether Caputo cheated by "intelligent design" or by "intelligent chance," let us not quibble, but generously accept that the explanatory filter reaches the same conclusion as the test: Caputo cheated. The shortcomings of the filter are nevertheless obvious, even in such a simple example.

In *No Free Lunch*, Dembski attempts to apply the filter to a real biological problem: the evolution of the bacterial flagellum, the little whip-like motility device some bacteria

such as *E. coli* possess. Dembski discusses the number and types of proteins needed to form the different parts of the flagellum and computes the probability that a random configuration will produce the flagellum (using the analogy of shopping randomly for cake ingredients). He concludes it is so extremely improbable to get anything useful that design must be inferred.

A comparison of Dembski's treatments of the Caputo case and the flagellum is highly illustrative, focusing on two aspects. First, in each case, Dembski only considers one chance hypothesis—the uniform distribution over possible sequences and protein configurations, respectively. He presents no argument as to why rejecting the uniform distribution rules out every other chance hypothesis. Instead, he shifts the burden of proof to the “design skeptic,” who, according to Dembski, “needs to explicitly propose a new chance explanation and argue for its relevance.” In the Caputo case, it may be warranted to test only one chance hypothesis, as there is only one such hypothesis that equates to fairness, but the situation is radically different for the flagellum, where nonuniformity in no way contradicts an evolutionary process of mutation and natural selection. Dembski routinely uses the uniform distribution as a synonym for lack of knowledge, a dubious practice that has been gainfully exposed by probabilist Olle Häggström.

Second, the one specific sequence of Democrats and Republicans that Caputo produced must be put together with other comparable sequences to obtain the rejection region. More specifically, we need to consider the set of 42 sequences that have at least 40 Democrats and compute its probability. Dembski does this correctly in the Caputo case, but when it comes to the flagellum, he does not consider the rejection region; he simply computes the probability of the outcome.

Dembski's way around this problem is to use his own term, “specification,” a vague concept that does not have a strict mathematical definition, but is intended to be a generalization of rejection region. In an essay titled “Specification: The Pattern That Signifies Intelligence,” it is said that “Specification denotes the type of pattern that highly improbable events must exhibit before one is entitled to attribute them to intelligence.” In *No Free Lunch*, the index entry “Specification, definition of” leads to a page where specification is used as a synonym for rejection region. The filter requires us at some point to compute a probability, so whatever “specification” is, it must be possible to convert it into the mathematical object of a set.

In the Caputo case, the two descriptions are easily integrated, as cheating can be described as patterns of the type “more Ds than Rs,” which also correspond to sets of sequences. However, when it comes to biological applications such as the flagellum, Dembski merely claims specification “always refers to function” and develops it no further.

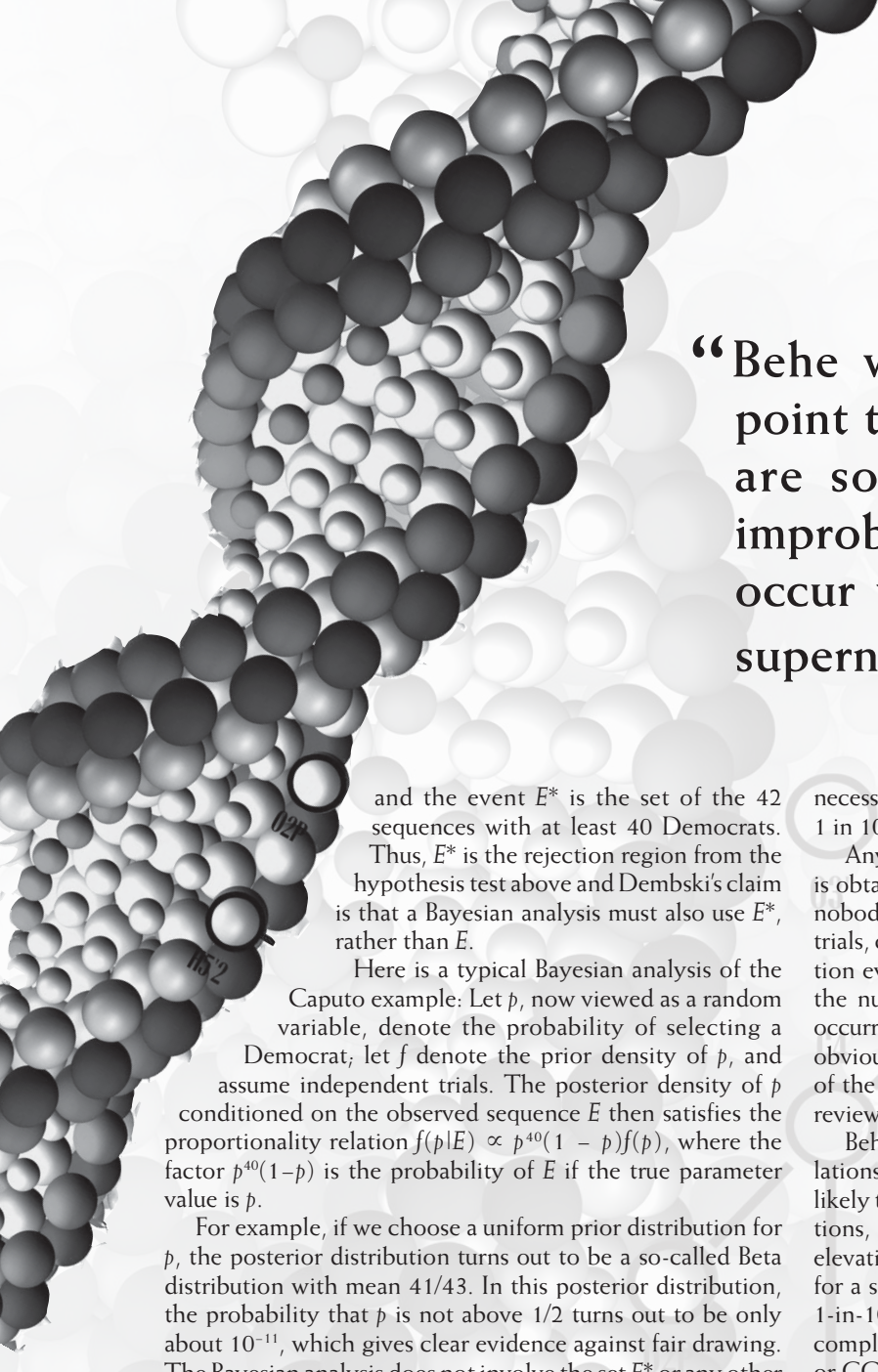
As opposed to the simple Caputo example, it is now very unclear how a relevant rejection region would be formed. The biological function under consideration is motility, and one should not just consider the exact structure of the flagellum and the proteins it comprises. Rather, one must form the set of all possible proteins and combinations thereof that could have led to some motility device through mutation and natural selection, which is, to say the least, a daunting task.

A general point of criticism against ID is that it does not offer any scientific explanations of natural phenomena, but merely attempts to discredit Darwinian evolution, aiming at inferring ‘design’ by default. Dembski's filter is streamlined to this approach; by trying to rule out all chance hypotheses, it attempts to infer design without stating any competing design hypotheses.

Above, it was demonstrated how the filter runs into trouble, even when it is viewed entirely within Dembski's chosen paradigm of “purely eliminative” hypothesis testing. Others have criticized the eliminative nature of the filter, claiming that useful design inference must be comparative. In a chapter titled “Design by Elimination vs. Design by Comparison” in his book *The Design Revolution*, Dembski counters this type of criticism. He starts by doing a ‘reality check’ to conclude that “the sciences look to Ronald Fisher and not Thomas Bayes for their statistical methodology,” referring to the divide in the statistical community (to the extent that such a divide really exists) between the frequentist approach—in which unknown parameters are viewed as constants and are subject to hypothesis testing—and the Bayesian approach—in which unknown parameters are viewed as random variables described by their probability distributions. However, the type of pure elimination he devises is not how statistical hypothesis testing is done in the sciences. A null hypothesis H_0 is not merely rejected; it is rejected in favor of an alternative hypothesis H_A . Moreover, one can compute the likelihood of the data for various parameter choices specified by H_A to conclude the evidence is, indeed, in favor of H_A (so-called power calculations). Hence, the statistical methodology of the sciences is eliminative and comparative.

One reason for Dembski to try to align with the frequentist camp is that there are indisputable problems with “Bayesian design inference.” For example, to apply Bayesian methods, one would have to assign a prior probability distribution over various chance and design hypotheses, which is obviously a more or less hopeless task. Dembski is not satisfied with such limited countercriticism, but decides to take on Bayesian inference altogether. In doing so, he claims Bayesian inference is “parasitic on the Fisherian approach,” as a Bayesian analysis must also use rejection regions! He even claims Bayesians do so “routinely,” but does not offer any examples. As the entire Bayesian approach is completely incompatible with the concept of hypothesis testing in general and rejection regions in particular, any such example would surely rock the world of statistics.

To illustrate his point, Dembski instead revisits the Caputo example. In his notation, the event E is the observed sequence of 40 Democrats and one Republican in some fixed order,



“Behe wishes to make the valid point that microbial populations are so large that even highly improbable events are likely to occur without the need for any supernatural explanations...”

and the event E^* is the set of the 42 sequences with at least 40 Democrats. Thus, E^* is the rejection region from the hypothesis test above and Dembski's claim is that a Bayesian analysis must also use E^* , rather than E .

Here is a typical Bayesian analysis of the Caputo example: Let p , now viewed as a random variable, denote the probability of selecting a Democrat; let f denote the prior density of p , and assume independent trials. The posterior density of p conditioned on the observed sequence E then satisfies the proportionality relation $f(p|E) \propto p^{40}(1-p)f(p)$, where the factor $p^{40}(1-p)$ is the probability of E if the true parameter value is p .

For example, if we choose a uniform prior distribution for p , the posterior distribution turns out to be a so-called Beta distribution with mean 41/43. In this posterior distribution, the probability that p is not above 1/2 turns out to be only about 10^{-11} , which gives clear evidence against fair drawing. The Bayesian analysis does not involve the set E^* or any other rejection regions. To do Bayesian design inference, one would need to augment the parameter space to allow for various design hypotheses and compute their respective likelihoods. Regardless of how this would be done practically, no rejection regions would ever be formed.

An ID Probability Challenge to Evolution

Michael Behe has presented his criticism of evolutionary biology in two books: *Darwin's Black Box*, published in 1996, and *The Edge of Evolution*, the 2007 follow up. The former does not contain much mathematics, but, in *The Edge of Evolution*, Behe has a chapter titled The Mathematical Limits of Darwinism, where he attempts to use probability and statistics to argue the case for ID.

Behe's central argument against human evolution hinges on how the malaria parasite *P. falciparum* has become resistant to chloroquine. The reason for invoking the malaria parasite is an estimate from the literature that the set of mutations

necessary for chloroquine resistance has a probability of about 1 in 10^{20} of occurring spontaneously.

Any statistician is bound to wonder how such an estimate is obtained, and, needless to say, it is very crude. Obviously, nobody has performed huge numbers of controlled binomial trials, counting the numbers of parasites and successful mutation events. Rather, the estimate is obtained by considering the number of times chloroquine resistance has not only occurred, but taken over local populations—an approach that obviously leads to an underestimate of unknown magnitude of the actual mutation rate, according to Nicholas Matzke's review in *Trends in Ecology & Evolution*.

Behe wishes to make the valid point that microbial populations are so large that even highly improbable events are likely to occur without the need for any supernatural explanations, but his fixation on such an uncertain estimate and its elevation to paradigmatic status seems like an odd practice for a scientist. Behe states a definition that incorporates the 1-in- 10^{20} figure: "Let's dub mutation clusters of that degree of complexity—1 in 10^{20} —'chloroquine-complexity clusters,' or CCCs."

He then goes on to claim that, in the human population of the last 10 million years, where there have only been about 10^{12} individuals, the odds are solidly against such an unlikely event occurring even once. In Behe's own words and italics:

On average, for humans to achieve a mutation like this by chance, we would need to wait a 100 million times 10 million years. Since that is many times the age of the universe, it's reasonable to conclude the following: *No mutation that is of the same complexity as chloroquine resistance in malaria arose by Darwinian evolution in the line leading to humans in the past 10 million years.*

On the surface, his argument may sound convincing. We humans are tremendously complex, and the malaria parasite consists of only one cell. Clearly, it would be absurd to claim we have evolved without experiencing even one mutation as complex as the little bug demonstrably has done. But one does not have to scratch deeply below the surface to recognize problems with Behe's statements.

First, he leaves the concept "complexity" undefined—a practice that is clearly anathema in any mathematical analysis. Thus, when he defines a CCC as something that has a certain "degree of complexity," we do not know of what we are measuring the degree. Lack of a clear definition is a fundamental problem when asserting something is proved, but let us nevertheless look further at Behe's claims.

As stated, his conclusion about humans is, of course, flat out wrong, as he claims no mutation event (as opposed to some specific mutation event) of probability 1 in 10^{20} can occur in a population of 10^{12} individuals (an error similar to claiming that most likely nobody will win the lottery because each individual is highly unlikely to win). Obviously, Behe intends to consider mutations that are not just very rare, but also useful, as can be concluded from his statement, "So, a CCC isn't just the odds of a particular protein getting the right mutations; it's the probability of an effective cluster of mutations arising in an entire organism."

Note that Behe now claims CCC is a probability; whereas, it was previously defined as a mutation cluster, another confusion arising from Behe's failure to give a precise definition of his key concept.

A problem Behe faces is that "rarity" can be defined and ordered in terms of probabilities; whereas, he suggests no separate definition of "effectiveness." For an interesting example, also covered by Behe, consider another malaria drug, atovaquone, to which the parasite has developed resistance. The estimated probability is here about 1 in 10^{12} , thus a much easier task than chloroquine resistance. Should we then conclude atovaquone resistance is a 100 million times worse, less useful, and less effective than chloroquine resistance? According to Behe's logic, we should.

Behe makes a point of his probability of 1 in 10^{20} being estimated from data, rather than calculated from theoretical assumptions. This approach leads to a catch-22 situation if we consider the human population with its 10^{12} members. Behe's claim is that there has not been a single CCC in the human population, and thus Darwinian evolution is impossible. But, if a CCC is an observed relative frequency, how could there possibly have been one in the human population? As soon as a mutation has been observed, regardless of how useful it is to us, it gets an observed relative frequency of at least 1 in 10^{12} and is thus very far from acquiring the magic CCC status. Think about it. Not even a Neanderthal mutated into a rocket scientist would be good enough; the poor sod would still decisively lose out to the malaria bug and its CCC, as would almost any mutation in almost any population.

In the above sense, Behe's claim is vacuously true. On the other hand, Behe has now painted himself into a corner, where he cannot obtain any empirical evidence for design because, as soon as a mutation has been observed, its existence is attributable to Darwinian evolution by population number arguments alone. Does there exist any population of any species where some individuals carry a useful mutation and others do not, such that this mutation can be explained by Darwinian evolution? Behe has already told us that one such example is chloroquine resistance in malaria. Does there exist any population of any species where some individuals carry a useful mutation and others do not, such that this mutation cannot be explained by Darwinian evolution? No. If one of n individuals experiences a mutation, the estimated mutation

probability is $1/n$. Regardless of how small this number is, the mutation is easily attributed to chance because there are n individuals to try. Any argument for design based on estimated mutation probabilities must therefore be purely speculative.

Arguments against the theory of evolution come in many forms, but most share the notion of improbability, perhaps most famously expressed in British astronomer Fred Hoyle's assertion that the random emergence of a cell is as likely as a Boeing 747 being created by a tornado sweeping through a junkyard. Probability and statistics are well developed disciplines with wide applicability to many branches of science, and it is not surprising that elaborate probabilistic arguments against evolution have been attempted. Careful evaluation of these arguments, however, reveals their inadequacies. ■

Further Reading

Elsberry, W. and Shallit, J. (2004) "Playing Games with Probability: Dembski's Complex Specified Information." In *Why Intelligent Design Fails: A Scientific Critique of the New Creationism*, Rutgers University Press: Piscataway, New Jersey.

Häggström, O. (2007) "Intelligent Design and the NFL Theorems." *Biology and Philosophy*, 22:217–230.

Matzke, N. (2007) "Book Review: The Edge of Creationism." *Trends in Ecology & Evolution*, 22:566–567.

National Academy of Sciences and Institute of Medicine (2008). *Science, Evolution, and Creationism*. Washington, DC: The National Academies Press. www.nap.edu/catalog/11876.html

Olofsson, P. (2008) "Intelligent Design and Mathematical Statistics: A Troubled Alliance." *Biology and Philosophy* (in press).

Perakh, M. (2003) *Unintelligent Design*. Prometheus Books: Amherst, New York.

Shallit, J. (2002) "Book Review: No Free Lunch." *BioSystems*, 66:93–99.

Sober, E. (2002) "Intelligent Design and Probability Reasoning." *International Journal for the Philosophy of Religion*, 52:65–80.