Biol Philos (2008) 23:545–553
DOI 10.1007/s10539-007-9078-6

Intelligent design and mathematical statistics: a troubled alliance

Peter Olofsson

Received: 28 February 2007 / Accepted: 4 June 2007 / Published online: 7 August 2007

Abstract The *explanatory filter* is a proposed method to detect design in nature with the aim of refuting Darwinian evolution. The explanatory filter borrows its logical structure from the theory of statistical hypothesis testing but we argue that, when viewed within this context, the filter runs into serious trouble in any interesting biological application. Although the explanatory filter has been extensively criticized from many angles, we present the first rigorous criticism based on the theory of mathematical statistics.

Keywords Intelligent design · Evolution · Mathematical statistics · Hypothesis testing

Introduction

A classic creationist argument against Darwinian evolution is that it is as likely as a tornado in a junkyard creating a Boeing 747. In recent years, the criticism has become more measured, coming not from young-earth creationists but from proponents of *Intelligent Design* (ID). The main claim of the ID proponents is that some biological phenomena are impossible to adequately explain without referring to design. The perhaps most prominent representative for the ID movement is biochemist Michael Behe whose 1996 book *Darwin's Black Box* (Behe 1996) presents challenges to Darwinian evolution based on *irreducibly complex* biochemical systems. A system is irreducibly complex if it consists of several different parts that are such that if any of them is removed, the system loses its function altogether. The favorite biological example is the *bacterial flagellum*, the little "outboard motor" that some bacteria are equipped with and to which we will return later. Behe's point is that Darwinian evolution cannot account for the emergence of irreducibly complex systems as all the parts need to be in place at once. Other than that, Behe seems to

P. Olofsson (✉)
Mathematics Department, Trinity University, One Trinity Place, San Antonio, TX 78212-7200, USA
e-mail: polofsson@trinity.edu

accept most of the accounts of Darwinian evolution. His claims in *Darwin's Black Box* have been thoroughly opposed, one of the most prominent critics being biologist Kenneth Miller (2000, 2004).

Whether the arguments against Darwinian evolution are based on tornadoes in junkyards or bacteria, the key concept for evolution critics is *improbability*. Since mathematics, probability, and statistics are highly developed disciplines, and are well established as indispensable scientific tools, it is only natural that evolution criticism has turned mathematical, trying to establish objective criteria to rule out chance explanations. The chief advocate for this approach is William Dembski whose ideas are described in his books *The Design Inference* (Dembski 1998) and *No Free Lunch* (Dembski 2002), and also in various postings on his own website (http://www.designinference.com).

In *The Design Inference*, Dembski introduces the *explanatory filter* as a generic method to eliminate chance explanations and infer design. Inspired by principles from statistical hypothesis testing, the explanatory filter aims at ruling out chance explanations of observed phenomena based on calculating their probabilities and argues that these probabilities are so small that chance is all but impossible. The filter is further discussed in *No Free Lunch* although in much less detail as the focus is this book is on mathematical complexity theory and the alleged failure of evolutionary algorithms (the title refers to a class of mathematical theorems that say, in essence, that in the absence of knowledge of the location of a target, nothing can beat blind search).

Dembski's oeuvre has been attacked from many different angles. He has had to endure criticism from biologists, philosophers, and mathematicians and the criticism spans the range from accusations of quasi-philosophy to discovery of arithmetic errors. Much of the criticism has been aimed at Dembski's forages into mathematical complexity theory and optimization theory, most notably by Jeffrey Shallit and Wesley Elsberry (Elsberry and Shallit 2003, 2004; Shallit 2002). An admirably short and surgically precise criticism of Dembski's use of the no-free-lunch theorems is presented by Häggström (2007).

The explanatory filter from *The Design Inference* has also been extensively criticized, perhaps most notably by philosopher Elliot Sober whose articles (Fitelson et al. 1999; Sober 2002, 2004), amongst many other things, criticize its purely eliminative nature, advocating instead that sound scientific practice require that conclusions are based on comparative reasoning. Perhaps a chance hypothesis confers a small probability on the evidence, but how do we know that a design hypothesis does not confer an even smaller probability? As the strategy of the ID movement is to try to discredit Darwinian evolution without offering any substantive alternative theory, the question is not easily answered. In *No Free Lunch*, Dembski argues against Sober that elimination is indeed a legitimate scientific principle. As an example, Dembski offers the hypothesis that the moon is made of cheese which he claims could be rejected without suggesting an alternative lunar material. The filter has also been criticized by Mark Perakh (2003, 2005) and by several others, for example on the websites http://www.talkreason.org and http://www.pandasthumb.org.

However, there has not yet been published a criticism of the filter from the vantage point of mathematical statistics. It is clear from Dembski's writings that his main source of inspiration for the explanatory filter is the theory of statistical hypothesis testing and a mathematical statistician immediately recognizes it as such. We shall see that, from this particular angle, even if the filter is put in the most benevolent light, it runs into serious trouble when it comes to biological applications. The filter will be described in the next section; thereafter we will outline how it relates to statistical hypothesis testing and finally address its weaknesses.

The explanatory filter

Dembski's flowchart description of the explanatory filter differs slightly between *The Design Inference* and *No Free Lunch*, but the basic structure is the same. An event is observed and it must be decided whether it is attributable to *regularity*, *chance*, or *design*. The first category refers to events that are bound to occur by natural law or are extremely likely to occur, such as managing to inhale an oxygen molecule when you draw a breath.

The second category refers to events that are neither certain nor extremely likely, but still occur without the need for any explanation beyond contingency. Obvious examples are coin tosses, die rolls, and spins of roulette wheels where chance reigns supreme. It is when probabilities are really minuscule that chance starts becoming questionable but a small probability alone is still not enough to rule out chance if the occurrence can be attributed to necessity or repetition. For an example of necessity, suppose that you are dealt a bridge hand. Any particular such hand is very improbable (a little over one in a trillion) but as you must always get *some* hand, you will always observe a one-in-a-trillion outcome. For an example of repetition, consider a state lottery. Any particular ticket is highly unlikely to win, but considering how many tickets that are bought, it is all but certain that eventually somebody wins, somewhere.

What, then, characterize the events that can not be explained by chance but must finally be attributed to design? First, they must of course have a small probability, or in Dembski's terminology, they must be *complex*.[1] Dembski explores various criteria for what exactly is a small enough probability, and although these criteria are debatable, the use of probability considerations *per se* is hardly controversial so let us simply adopt the working convention that we could all agree on some critical probability value. Second, as illustrated by the examples above, it must be possible to rule out the possibility that the events occur by necessity or repetition. It is intuitively clear what type of events these would be. For example, when you are dealt the bridge hand, you would be highly surprised to get thirteen hearts even though it has the same one-in-a-trillion probability as any other hand, and would probably suspect that the deck had been stacked. The particular feature of the hand of thirteen hearts is, in Dembski's terminology, that it is *specified* and this makes you think that something other than chance was responsible.

The events that are finally assigned to the design category are thus those that are complex and specified. It should be noted here that nothing is said about the identity or purpose of the designer; the concept "design" is simply taken to mean "neither regularity, nor chance." Dembski spends a lot of time and effort trying to establish a mathematical definition of *specified complexity*, using concepts from probability theory and mathematical complexity theory. However, when it comes to the biological applications that are at the very heart of the matter, Dembski simply claims that any biological system that has a recognizable function must be specified, and also adds that no biologist he knows would question this conclusion. In their review of *No Free Lunch*, Elsberry and Shallit (2003, 2004) point out the obvious equivocation in Dembski's use of "specified," first as a technical concept with a strict definition and then in its everyday meaning. Considering how incredibly complicated biological systems are, establishing a definition that is both mathematically tractable and biologically applicable is a daunting task and it is probably fair to say that the final solution has not yet arrived.

---

[1] This use of the term "complex" is at odds with its usual meaning in mathematics, see (Elsberry and Shallit 2003, 2004; Perakh 2005).

Let us leave the mathematical problems aside and accept that biological systems are eligible to be investigated with the explanatory filter. In the next section, we will examine the filter in more detail and also describe how it relates to statistical hypothesis testing.

The filter and statistical hypothesis testing

In one way, the explanatory filter appeals to mathematicians and statisticians. After all, the structure of the explanatory filter is essentially that of mathematical proof by contradiction or, even more poignantly, that of statistical hypothesis testing. In fact, the final step in the explanatory filter *is* a statistical test but whereas statistical testing in general only rules our a particular chance hypothesis, the aim of the filter is to rule out chance altogether.

Let us look at one of Dembski's favorite examples, that of Nicholas Caputo, a New Jersey Democrat who was in charge of putting together the ballots in his county. Supposedly, there is an advantage in having the top line of the ballot, and Caputo managed to place a Democrat on the top line in 40 out of 41 elections. Republicans filed a lawsuit and the case ended up in the New Jersey Supreme Court. The crucial argument against Caputo was a probability calculation finding that a fair drawing procedure would produce such an extreme outcome with a probability of less than 1 in 50 billion. Based on the extreme odds against Caputo's ballot lines, the Court suggested that Caputo institute new guidelines for ballot line selections. Where does the 1-in-50-billion probability come from and what does it have to do with the explanatory filter?

Before we apply the explanatory filter to Caputo's sequence of ballot lines, let us first agree that Caputo's sequence is specified. In accordance with the discussion of biological specificity in the previous section, let us not argue over the precise mathematical meaning but simply note that somebody who is a Democrat, selects ballot lines without public oversight, and has a last name that sounds like a combination of *Capone* and *Bluto* might have an interest in favoring Democrats. Caputo's sequence is specified in the sense of indicating a pattern of cheating.

Passing the first step of the filter requires us to rule out the possibility that Caputo inadvertently used a flawed randomization method. To obtain a sequence of Ds and Rs in a fair manner, the sequence must be obtained by a randomization device that each time makes D and R equally likely and is such that the outcome each year is independent of previous years' outcomes. Tossing a fair coin would achieve this. If Caputo instead used a random number generator, this generator might have been flawed in the sense of generating Ds much more often than Rs. It could also have violated the independence requirement by having a strong tendency to repeat its last outcome, only occasionally changing between D and R. By Caputo's own account, he had used a fair procedure by drawing capsules from urns and by taking his word, we rule out regularity and proceed to the next step in the filter.

To pass the second step, we must compute a probability and decide that it is small enough to rule out chance. The question is which probability to compute. Suppose that Caputo's sequence looked like this:

DDDDDDDDDDDDDDDDDDDDDD**R**DDDDDDDDDDDDDDDDDD

Under the assumption of a fair drawing procedure, this sequence has probability $(1/2)^{41}$ which is less than 1 in 2 *trillion*. However, *any* sequence of 41 Ds and Rs has the same probability so this number alone is not enough evidence. Remember that we have decided that Caputo's sequence is specified in the sense that it indicates a pattern of cheating. Still, Caputo's sequence does not represent the only way in which he could have cheated. Any of the 41 sequences with 1 R and 40 Ds would be equally indicative of cheating and the sequence with 41 Ds and no Rs at all would be even more so. What we need to do is to consider the set of all sequences that are *at least* as extreme as Caputo's sequence and compute the probability of this set. The probability that a fair procedure produces a sequence with at least 40 Ds turns out to be less than 1 in 50 billion which provides strong evidence against Caputo. The set of extreme sequences is known to statisticians as a *rejection region*; any outcome in this region leads to a rejection of the hypothesis of fair drawing.[2] Dembski introduces the term *specification* to denote rejection regions. A specification is thus the set of all specified outcomes that carry evidence against a hypothesis in at least as high a degree as the observed outcome.

It is important to note that it is the probability of the rejection region, not of the individual outcome, that warrants rejection of a hypothesis. A sequence consisting of 22 Ds and 19 Rs could also be said to exhibit evidence of cheating in favor of Democrats, and any particular such sequence also has less than a 1-in-2-trillion probability. However, when the relevant rejection region consisting of all sequences with at least 22 Ds is created, this region turns out to have a probability of about 38% and is thus easily attributed to chance.

Now that Caputo's sequence has passed the second step, the explanatory filter rules out regularity and chance, and infers design (in this case, cheating). In contrast, a statistical hypothesis test of the data would typically start by making a few assumptions, thus establishing a *model*. If presented with Caputo's sequence and asked whether it is likely to have been produced by a fair drawing procedure, a statistician would first assume that the sequence was obtained by each time independently choosing D or R, such that D has an unknown probability p and R has probability $1-p$. The statistician would then form the *null hypothesis* that $p = 1/2$ which is the hypothesis of fairness. In this case, Caputo would be suspected of cheating in favor of Democrats so the *alternative hypothesis* would be that $p > 1/2$, indicating that Ds were more likely to be chosen. Next, it would be noted that the rejection region of at least 40 Ds has a very small probability, and the null hypothesis of fairness would be rejected in favor of the alternative hypothesis.

The difference between the explanatory filter and the hypothesis test is subtle. The filter started by ruling out, based on Caputo's own account, the alternative hypothesis, and then tested the only remaining chance hypothesis, that $p = 1/2$. Once this final hypothesis is rejected, nothing remains but to infer design. In contrast, the hypothesis test started directly at the second step, only rejecting the particular hypothesis that $p = 1/2$. Upon hearing Caputo's account, however, the statistician will realize that his model assumptions were incorrect and in the end reach the same conclusion as the "design theorist." Indeed, it is a general observation that an unlikely outcome may not only cast doubt on the null hypothesis but on the entire statistical model. In this regard, the explanatory filter is merely a description of the entire procedure of choice of model and hypothesis test.

Criticism of the filter

Thus far, we have not voiced any serious objections to the explanatory filter. Viewed on a sufficiently abstract level, it does appear to be logical and based on well-established principles and techniques from mathematical statistics. But no more Mr Nice Guy. The

---

[2] Technically, a rejection region is set in advance and corresponds to a prespecified probability called the *significance level*. The 1-in-50-billion probability that shows up here is what statisticians call a *p-value*.

filter runs into serious trouble when it comes to biological applications and nothing illustrates this better than Dembski's chosen example of the flagellum of the bacterium *E. coli*.

In *No Free Lunch*, Dembski sets out to run the bacterial flagellum through the explanatory filter. He 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. He concludes that it is so extremely improbable to get anything useful that design must be inferred. He admits that the assumptions and computations are simplified but as such simplifications arise in any mathematical model, we will not hold it against him.

My first objection is that Dembski *considers only the outcome, not the rejection region*. Recall that we need to consider not only what *is* but also what *could be*. In the Caputo example, the observed outcome was combined with other potential outcomes carrying as much, or more, evidence against a fair drawing procedure. It is the probability of this set that is crucial, not the probability of the observed outcome. For the flagellum, the outcome is what can be observed, namely, the particular combination of proteins that form the different parts. But what is the rejection region?

Dembski's way around this problem is to ignore it and mask this by a subtle equivocation. Recall that he introduced the term "specification" to denote a rejection region, which is a set of outcomes with a certain common property. When it comes to biological systems, Dembski states that "specification always refers to function." In the case of the bacterial flagellum, specification would thus refer to motility, and the flagellum itself represents a specified outcome in the set of all possible mechanisms that achieve motility as well, or better. However, in his application to the flagellum, Dembski does not even discuss specification but seems to equate it with the flagellum itself (which is "specified" but not a "specification"). It is understandable though, because how could he possibly address what it means to have "at least the function of the flagellum?" This task would include suggesting a variety of possible genetic make-ups that would lead to a motility device of equal or greater efficiency than the flagellum.[3]

Presumably, Dembski intends his concept of specification to be more general than what a statistician means by a rejection region. In *The Design Inference*, there is a mathematical definition of specification, and in *No Free Lunch*, specification is introduced as a rejection region (it is not stated as a definition but it is the page in the book where one is referred by the index entry "Specification, definition of"). Dembski's later writings indicate that defining specification is still a work in progress, but regardless, the problem of potential but unknown scenarios will always be there, for the flagellum as well as for any other biological system. Bacterial propellers are so much more complicated than cheating New Jersey politicians.

The second problem with Dembski's application is that he *only tests one particular chance hypothesis* and takes its rejection as evidence that all chance hypotheses can be ruled out. In particular, the chance hypothesis he considers is based on a uniform distribution or, in daily language, that the flagellum is assembled at random. Once more, Dembski fails to apply his filter as he has described it. Remember that he should rule out regularity before he goes on to compute probabilities. Granted, there is no known natural law that would automatically assemble the flagellum but he also needs to rule out other chance explanations. An evolutionary biologist would certainly argue that, according to

---

[3] Dembski's calculation of "perturbation probabilities" is something else as this refers to how much variation that is allowed in the protein assembly and still retain the basic function of the flagellum.

some plausible evolutionary scenario, the formation of the flagellum is an event of a probability that is far from negligible. Dembski does not address this possibility but starts directly at the final step of the filter.

There is an important point to be made here. The probability of an event depends on what chance hypothesis, or *probability distribution*, that is operating. For a quick and common example, consider the Shakespearean phrase TO BE OR NOT TO BE. If 13 letters are chosen at random, what is the probability to get this phrase? The everyday meaning of "chance" and "at random" is that letters are chosen independently and that all letters of the alphabet are equally likely to be chosen. However, in probability theory this is merely one example of a chance hypothesis corresponding to what is know as a *uniform* probability distribution. Using this distribution, it is very unlikely to get the phrase but there are many other plausible probability distributions that confer different probabilities on the phrase. For example, if letters are chosen according to their frequencies in the English language, the phrase becomes more probable. If, in addition, a letter is chosen in accordance with how likely it is to follow another letter, the phrase becomes even more likely (in this case, letters are not chosen independently of each other). Any time a probability distribution (and dependence structure) is specified, the probability of the phrase can be computed, and can be anything between 0 and 1.

Let us return to the flagellum where Dembski considers only the uniform distribution, thus assuming that all protein configurations are equally likely. This is yet another version of the old creationist classic: a microscopic tornado in a protein junkyard (although Dembski's own allegory is to go on a random shopping spree for cake ingredients). In a sense, Dembski achieves absolutely nothing as no evolutionary biologist would suggest a model even remotely resembling Dembski's shopping cart and would thus gladly agree to rule out this particular chance hypothesis. But whereas the evolutionary biologist would have in mind a more realistic chance scenario, Dembski rules out chance altogether. His argument for doing so is discussed in a section of *No Free Lunch* where he states the need for "sweeping the field of chance hypotheses." Writes Dembski:

> Design inferences therefore eliminate chance in the global sense of closing the door to all relevant chance explanations. To be sure, this cannot be done with absolute finality since there is always the possibility that some crucial probability distribution was missed. Nonetheless, it is not enough for the design skeptic merely to note that adding a new chance explanation to the mix can upset a design inference. Instead, the design skeptic needs to explicitly propose a new chance explanation and argue for its relevance for the case at hand.
>
> *No Free Lunch*, p. 67–68

Thus, once Dembski has ruled out a chance hypothesis of his choice, the burden of proof is on the "design skeptic" who must suggest a relevant chance hypothesis and also compute its probability. This is in stark contrast to Dembski's argument for pure elimination in a later section:

> What's more, a proposed solution may be so poor and unacceptable that it can rightly be eliminated without proposing an alternative (e.g., the moon-is-made-of-cheese hypothesis). It is not a requirement of logic that eliminating a hypothesis means superseding it.
>
> *No Free Lunch*, p. 102

The inconsistency is striking. Design skeptics are required to make sure that a rejected hypothesis is superseded; design proponents are not. Without doubt, most biologists would consider Dembski's shopping cart model "poor and unacceptable" and thus, appealing to logic, could safely eliminate it and go on to more important activities.

Concluding remarks

The explanatory filter may be logically sound but it is virtually impossible to apply. The criticism presented against Dembski's application to the bacterial flagellum may be countered by pointing out that this is merely one example that is far from complete, which Dembski also readily acknowledges. In addition, he has no monopoly on the filter and there may be those who are more successful in applying it than Dembski himself. Regardless of any such objections, the application to the flagellum exemplifies the problems that will show up in any application to any even mildly complicated biological system.

The first problem, how to describe a relevant rejection region seems, if at all possible, like an especially daunting task. Suggestions anyone?

The second problem, unrealistic use of the uniform distribution, is at least possible to discuss. The probability of a biological system under assumptions of the uniform distribution might often be possible to compute, at least after some simplifying model assumptions. This probability will often be extremely small, and even though one could argue that the assumptions are deeply flawed to start with and that Dembski has no right to demand an alternative hypothesis, it is still the nature of scientists to assume the burden of proof. Biologists are interested in understanding the bacterial flagellum, and realize that it may pose huge challenges, irrespective of any criticism from the ID community. However, it is unreasonable to demand that the biologist be able to suggest a hypothesis that admits probability calculations. It is one thing to compute probabilities assuming unrealistic random assembly, quite another to compute probabilities based on realistic models of millions of years of evolution, reproduction, and natural selection. In the words of prominent mathematical biologist Martin Nowak (2005), "We cannot calculate the probability that an eye came about. We don't have the information to make this calculation."

In a way, the ideas in *The Design Inference* and *No Free Lunch* are examples of an exaggerated belief in mathematical methods in the sciences. Mathematical methods are of course extremely useful, but not equally so in each scientific discipline. Theoretical physics, for example, would not exist without mathematics, but the situation is quite different in biology where systems and processes are much more complicated (from a mathematical point of view). Mathematics, probability, and statistics can be, and have been, very successfully applied in many fields of biology. However, there are also many obstacles and limitations and as we have seen, these are alarmingly present in attempted applications of the explanatory filter.

References

Behe M (1996) Darwin's black box. Free Press, New York

Dembski WA (1998) The design inference: eliminating chance through small probabilities. Cambridge University Press, Cambridge

Dembski WA (2002) No free lunch: why specified complexity cannot be purchased without intelligence. Roman & Littlefield, Lanham, MA

Elsberry W, Shallit J (2003) Information theory, evolutionary computation, and Dembski's complex specified information. Submitted for print; also available at http://www.talkreason.org/articles/eands-dembski.pdf

Elsberry W, Shallit J (2004) Playing games with probability: Dembski's complex specified information, In: Young M, Edis T (eds) Why intelligent design fails: a scientific critique of the new creationism. Rutgers University Press, New Brunswick, pp 121–138

Fitelson B, Stephens C, Sober E (1999) How not to detect design—a review of William Dembski's the design inference. Philos Sci 66:472–488

Häggström O (2007) Intelligent design and the NFL theorems. Biol Philos 22:217–230

Miller K (2000) Finding Darwin's god: a scientist's search for common ground between God and evolution. Harper Perennial, New York

Miller K (2004) The flagellum unspun: the collapse of irreducible complexity, In: Dembski WA, Ruse M (eds) Debating design: from Darwin to DNA. Cambridge University Press, Cambridge

Nowak M (2005) Time magazine, August 7, 2005

Perakh M (2003) Unintelligent design. Prometheus Books, New York, 459 pp

Perakh M(2005) The dream world of William Dembski's creationism. Skeptic 11(4)

Shallit J (2002) Book review: no free lunch. BioSystems 66:93–99

Sober E (2002) Intelligent design and probability reasoning. Int J Philos Relig 52:65–80

Sober E (2004) The design argument. In: Mann W (eds) The blackwell guide to philosophy of religion, Blackwell, Oxford, pp 117–147