**Inference to the Best Explanation and Dembski-Significance Testing Model for the Intelligent Design Argument**

**Chong Ho Yu, Ph.D.**

**Paper submitted to *Perspectives on Sciences and Christian Faith***

**RUNNING HEAD: IBE and DSTA**

**Correspondence:**

**Chong Ho Yu, Ph.D.**

**PO Box 612**

**Tempe AZ 85280 USA**

**Email: chonghoyu@yahoo.com**

**Abstract**

In the debate regarding the Design Inference (DI), philosopher of biology Sober frames the issue in terms of the Inference to the Best Explanation (IBE), and he develops his counterarguments against DI by applying the Fisherian Likelihood Principle. Dembski, a mathematician/philosopher who is also influenced by Fisher, constructs his pro-DI arguments by employing Fisherian significance testing and rejection region. In addition, based upon the bad lot argument, Dembski asserts that being forced to choose among competing explanations, as is suggested by IBE, is unacceptable. This article is a brief evaluation of the debate between Sober and Dembski. It is suggested that Dembski's application of rejection region may not be appropriate to DI when the cut-off level for the specification of the rejection region is controversial and the chance hypothesis is eliminated without taking other probability distributions into consideration. Moreover, the bad lot argument does not negate the value of abduction/IBE since abduction operates in an exploratory, not confirmatory mode, and IBE proposes using competing, not alternative explanations. It is recommended that the statistical and probabilistic approach for investigating Intelligent Design should be first framed in the abductive mode, which places the emphasis on hypothesis generation instead of elimination based upon a pre-determined distribution and rejection region. Further, it is argued that the Fisherian and the Bayesian approaches are not necessarily mutually exclusive; taking the Bayesian approach into account could strengthen, not weaken, the Intelligent Design argument.

**Inference to the Best Explanation and Dembski-Significance Testing Model for the Intelligent Design Argument**

In the debate regarding the Design Inference (DI), a hypothesis that the universe originated from an intelligent design, certain opponents and proponents assert that the structure of the argument is Inference to the Best Explanation (IBE), which is a form of abductive reasoning advocated by the American philosopher Charles Sanders Peirce and further developed later by Harman.[1] In recent years Dembski, a well-known Christian mathematician/philosopher, has been developing an alternative to IBE to argue for DI, but his method has been disputed by Sober, who is a prominent philosopher of biology.[2] Nonetheless, both Sober and Dembski build their arguments upon different methods of probabilistic inference. While Sober relies on the likelihood principle to argue against DI, Dembski develops his pro-DI argument on the basis of Fisherian significance testing and rejection region. Dembski explicitly distinguishes his methodology, “the Fisher approach,” from Sober’s “likelihood approach”, and Koons also refers to Dembski’s approach as the “Dembski/Fisher model”.[3] However, this may be a misnomer because both the rejection region and the likelihood principle were developed by R. A. Fisher. To avoid confusion, in this article I refer to Sober’s method as the “Sober-likelihood approach” (SLA) and Dembski’s as the “Dembski-significance testing approach” (DSTA). In my view both SLA and DSTA have certain epistemological and methodological difficulties. My criticisms of Sober can be found in Yu;[4] in this article I will concentrate on a discussion of DSTA. As a supporter of IBE, I found that DSTA seems to operate in a confirmatory rather than exploratory/discovery mode, and thus its appropriateness to a nonrepeatable event (the origin of life) is questionable. In addition, Dembski’s criticisms of SLA may be caused by certain misunderstandings of abduction and IBE.

## What is abduction/IBE?

IBE or abduction is widely applied in various disciplines, yet its interpretations are very diverse. As a result, many scholars question whether IBE or abduction is properly conceived. For example, the notion of abduction is popular in linguistics, but Deutscher argues that the linguistic abduction is based on a critical misunderstanding of Peircean philosophy.[5] Likewise, Rozeboom warns, "Abduction is at risk of becoming a buzzword in the Artificial Intelligence (AI) circle; and the extent to which psychology's research methods can profit study of AI data-processing algorithms claimed to be abductive is problematic." [6]

To compare and contrast DSTA and SLA, which is related to IBE, it is important to explain what abduction/IBE is. In Peirce's view, abduction is a form of reasoning that serves as a logic of discovery. To be specific, when a surprising phenomenon is observed, the researcher puts aside his/her predetermined conceptions and hypotheses; instead, new categories and concepts are formulated to cope with the surprising observation. Behrens and Yu illustrate this practice in the following story:

> After years of extensive field work an entomologist develops a prediction that butterflies with a certain spotting should exist on top of a particular mountain, and sets off for the mountain top. Clearly, if she finds such butterflies there will be evidence in support of her theory; otherwise, there is an absence of evidence. On her way to the mountain she traverses a jungle in which she encounters a previously unknown species of butterflies with quite unanticipated spottings. How does she handle this? Should she ignore the butterfly because she has not hypothesized it? Should she ignore it because it may simply be a misleading Type I error (the probability of rejecting the null hypothesis when the null is true)? Should she ignore it because she may change her original hypothesis to

> say she has really hypothesized this jungle butterfly? For most individuals it is clear that a new discovery is valuable and should be well documented and collected. Her failure to have hypothesized it does not impugn its uniqueness, and indeed many great scientific conclusions have started with unanticipated findings. Should she worry about Type I error? Since Type I error concerns long-run error in decision-making based on levels of specific cut-off values in specific distributions, that precise interpretation does not seem to matter much here. If she makes an inference about this finding then she should consider the probabilistic basis for such an inference, but nevertheless the butterfly should be collected. Lastly, should she be concerned that this finding will contaminate her original hypothesis? Clearly she should continue her travel and look for the evidence concerning her initial hypothesis on the mountaintop. If the new butterfly contradicts the existing hypothesis, then she has more data to deal with and additional complexity that should not be ignored. If she is concerned about changing her hypothesis in midstream to match the new data, then she has confused hypothesis generation and hypothesis testing. With regard to any new theories, she must create additional predictions to be tested in a different location.[7]

The moral of this story is that a surprising phenomenon or a new observation may not have a corresponding probabilistic distribution in the Fisherian sense, let alone a null hypothesis, an alternate hypothesis, and a Type I error rate. In the Fisherian tradition, probability is defined in the **long run**, which is comparable to inductive reasoning. However, abductive reasoning is concerned with the matter **at hand** rather than the long run. Many of Peirce's writings indicate that induction and abduction are two separate modes of reasoning.[8] As Hintikka says, "Peirce is one hundred percent right in denying the role of naked induction in forming new hypotheses" [9]

As mentioned before, the mission of Peircean abduction is to look for a pattern in a surprising phenomenon and then to suggest one or several plausible hypotheses. For example,

The surprising phenomenon, X, is observed.

Among hypotheses A, B, and C, A is capable of explaining X.

Hence, there is a reason to **pursue** A.

It is important to emphasize that in the Peircean framework, abduction is employed as a "question-answer step, not as an inference in any literal sense of the word."[10] At most, abduction could provide a conjecture or a potential hypothesis A to pursue, but it would not confirm or disconfirm A.

Nonetheless, when Harman extends the notion of abduction to IBE, IBE becomes a justification for adopting an explanation. In Harman's view, IBE leads us to adopt the most coherent and complete explanatory account that can fit into our **total explanatory picture** of the world while no competing hypothesis would do as well. In order to base the inference on this totality, IBE must be drawn upon **competing** explanations, not just **alternative** explanations.[11] In my interpretation, the competition in IBE must go beyond the conventional alternative-explanation approach, which has been widely employed in hypothesis testing. Nonetheless, sometime this so-called competition is an illusion. Yu illustrates this problem in the following example:

> What do you think if an engineer makes the following claim? 'This is a breakthrough in engineering science. Repeated experiments confirm that a Porsche 911 can outrun a Dodge Neon.' Look at the following two experiments: (a) Engineer A wants to test the engine performance of Porsche 911. He compares it against a Dodge Neon. (b) Engineer B wants to test the engine performance of Porsche 911. He compares it against a Ferrari.

> By common sense, most people will laugh at the first benchmark test and approve the second one. But look at the next pair: (a) Researcher A spent 100 hours to develop a Web-based course as the treatment. He simply printed out a hard copy of those web pages in half an hour for the control group. (b) Researcher B spent 100 hours to develop a Web-based course as the treatment. He also invested a lot of effort to develop a stand-alone *Director* (a multimedia authoring program) version for the control group. What Researcher A did is very similar to what Engineer A did. Very often educational research is a comparison between a Porsche 911 and a Dodge Neon.[12]

In classical hypothesis testing, the null hypothesis in the above "experiment" is that there is no significant statistical difference between Web-based instruction and paper-based instruction, which could be rejected easily. The alternative hypothesis is that somehow there is a difference. As Treitel observes, "If no plausible rival hypothesis is to be found, the scientist must manufacture or obtain at least one dummy rival hypothesis, and then formally show that it is inferior to the preferred method."[13] This so-called "competition" is totally legitimate in statistical and probabilistic inferences, but it is doubtful that this competition would lead to a conclusion carrying scientific merits. More importantly, this "competition" is set up with strong assumptions. For example, it is assumed that instructional method is the sole factor in performance difference. But this conceptualization could by no means capture the total explanatory picture of the world. In Harman's sense, competing explanations need not be competing explanations of the same thing. One could boldly hypothesize that family values, socioeconomic status, or religious belief is the key factor affecting school performance.

The preceding brief introduction to Peirce's idea of abduction and Harman's notion of IBE is by no means comprehensive, but this account can serve here as a starting point for evaluating the clash between SLA and DSTA.

**Fisherian legacy and Dembski-significance testing model**

Although Dembski does not explicitly label his support of DI as abduction or IBE, Dembski follows a path of reasoning that is similar to abduction: Life is specified and complex; this is a surprising phenomenon. There are three possible explanations: Regularity, chance, and agency. Dembski eliminates the first two and pursues the last one. He claims that his elimination approach follows the common statistical practice popularized by Fisher, which is to reject a chance hypothesis if a sample appears in a prespecified rejection region. Due to the law of small probability, life in the form of specified complexity does not seem to occur by chance and thus DI should be seriously considered.

Dembski explicitly supports the Fisherian school as the sole foundation of his probabilistic and statistical reasoning by downplaying the Bayesian School. He writes,

> The friendly (to the Design Inference) approach, due to Ronald Fisher, rejects a chance hypothesis provided sample data appear on a pre-specified region. The less friendly approach, due to Thomas Bayes, rejects a chance hypothesis provided an alternative hypothesis confers a bigger probability on the data in question than the original hypothesis…Whereas in the Fisherian approach the emphasis is on elimination, in the Bayesian approach the emphasis is on comparison…The sciences look to Fisher and not Bayes for their statistical methodology. Colin Howson and Peter Urbach, in *Scientific Reasoning: The Bayesian Approach*, likewise admit the underwhelming popularity of Bayesian methods among working scientists. [14]

No doubt Dembski is predominantly a Fisherian, but his ranking of the Fisherian school over the Bayesian is not widely agreed upon in the scientific community. As a matter of fact, today not only is the Bayesian approach widely accepted by the scientific community, but also the integration of various statistical schools of thought is popular.[15] Some researchers, such as Press and Tanur, bluntly present an anti-Fisherian and pro-Bayesian view:

> To many, it has become increasingly clear that the frequentist approach is fraught with technical problems and inconsistencies…As a consequence, today, scientists schooled in the Bayesian approach to scientific inference have been departing from the frequentist approach and returning to the Bayesian approach.[16]

It is debatable whether the Fisherian (frequentist) approach is "fraught with technical problems and inconsistencies," nonetheless, Press and Tanur's view is a counter-example to Dembksi's notion of "underwhelming popularity of Bayesian methods among working scientists." The issue of Bayesianism requires a paper on its own right, and thus it will be briefly discussed in the section "Discussion and recommendations" near the end. In the subsequent discussion the focus will be on IBE.

Before examining DSTA, it is essential to explain what the Fisherian school is. R. A. Fisher is a colorful figure in the history of statistics. Throughout his life he developed numerous statistical methodologies. Interestingly enough, on some occasions Fisher's followers in different camps adopted different notions of his to argue against each other. For example, in recent years the resampling school has promoted resampling methods such as randomization tests and bootstrapping, as opposed to the classical inference based upon theoretical probability distributions. Nevertheless, both randomization testing with empirical distributions and significance testing with theoretical distributions were developed by Fisher.[17] In the debate concerning DI, as mentioned before, both the likelihood and the rejection region approaches also originated from Fisher.

**Figure 1. Probability distribution and region of rejection.**

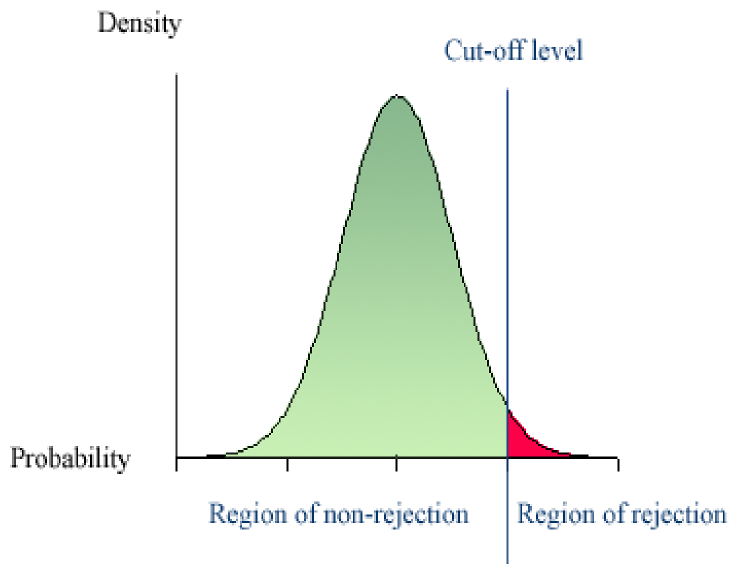


Figure 1 represents a theoretical probability distribution in classical Fisherian significance testing. The Y-axis denotes the probability value while the x-axis depicts the probability density. In the rejection region approach, an observed sample is compared against a theoretical probability distribution. In order to determine whether the occurrence of this sample is rare, a cut-off level, also known as the alpha level (the blue line) must be specified. An event falling outside the cut-off level (the green area) is not considered to have a small probability while an event falling inside the rejection region (the red area) leads to the rejection of the null hypothesis where there is no significance. In this context, the term “significance” is not equivalent to the term “importance” as it is commonly thought of; rather, it means the rarity of the event.[18]

It is important to note that in Fisherian significance testing there is no alternate hypothesis for comparison. Currently significance testing and hypothesis testing are used interchangeably, but indeed there is a subtle difference between the two. When Fisher introduced his methodology, there was only one hypothesis: the null hypothesis. Following this strategy, the only possible conclusions were to reject hypothesis or not. Later Neyman and Pearson introduced the concept of alternate hypothesis. The synthesis of the Fisherian school and the Neyman-Pearson school, in which at least two hypotheses are tested, has become a common practice of hypothesis testing. [19] This point is crucial because on the basis of rejection region and significance testing, which originated from Fisher rather than Neyman-Pearson, Dembski builds his argument against Sober and defends his case that DI is a viable explanation. In the following, I will raise several questions about his methodology and his counterargument against SLA.

## Questions on Dembski-significance testing approach

What probability distributions are applicable?

In the story told by Behrens and Yu about an entomologist who found a butterfly in an unexpected place, probabilistic modeling such as Type I error rate may not be applicable to that surprising phenomenon. Common questions when testing a chance hypothesis are: What is the probabilistic distribution that could be associated with the event? If the test result, based on a particular theoretical distribution, leads to the rejection of the chance hypothesis, is it possible that the result could be different under another distribution? Dembski provides the following answer:

> There is thus a crucial difference between the way statistics eliminates chance and the way the design inference eliminates chance. When statistics eliminates chance, it is always a particular probability distribution (or set of probability distributions) that gets

> eliminated, with the question remaining what alternative probability distributions might be operating in its place. On the other hand, when the design inference eliminates chance, it leaves no room for alternative probability distributions. Statistics, therefore, eliminates chance only in the limited sense of rejecting one chance explanation while leaving another open. The design inference, on the other hand, eliminates chance in the global sense of closing the door to every relevant chance explanation.[20]

However, I find this assertion unconvincing. It is important to note that performing Fisherian significance testing is to find out the probability that an event occurs in the long run by chance alone; this is based upon a frequentist interpretation of probability. However, the origin of life is a single and non-repeatable event, and thus there is no background information to determine which theoretical probability distribution is applicable. In order to estimate the probability of the occurrence of life and determine whether the probability is small enough to reject the chance hypothesis, the test needs to be conducted in a counterfactual manner rather than by collecting empirical samples. To be specific, there is no variance in the sample. There is no additional universe which has no life forms;, and also we could not find alternate life forms which are more or less specified than the known species. Thus, one could only permute combinations of particles or DNA sequences to simulate the probability of the occurrences of life in a “what-if” manner. However, this type of permutation is conducted without any background information.

Even if we grant that a significance test yields a small probability, one may ask the following questions: How small is small? Where should we draw the cut-off line? In classical hypothesis testing there are three common cut-off levels: 0.01, 0.05, 0.1. Among these three options, 0.05 is widely selected. However, this is an arbitrary cut-off level, as Rosnow and

Rosenthal point out when they quip, "Surely, God loves the .06 nearly as much as the .05." [21] The appropriateness of the cut-off depends on the situation. For instance, if a scholar conducts exploratory research, he/she may set a looser cut-off level in order to include more potential variables for further investigation. On the other hand, if a scientist operates the study in a confirmatory mode, the cut-off level may be as strict as 0.01. When the chance is extremely small, it could have real significance. Moreover, the meaning of significance or rarity is also tied to the nature of the discipline. The number alone indicates the statistical significance only. Determining the practical significance and the clinical significance is a matter of background information. To be specific, in an educational setting, a probability of 0.05 may not be “significant” enough for policy makers to alter the existing instructional practice, but in a medical setting a tiny increase or decrease in probability is a matter of life or death. However, in the context of DI, researchers have little or no background information from which to set the cut-off level and interpret the “significance” of the probability. Dembski proposes that the event is considered unlikely if the probability that such an event could occur once in the entire history of the physical universe is below one-half. Dembski may need to provide a stronger rationale for choosing this number and closing the case after a single test of the chance hypothesis.

Likelihood principle and competing explanations

Sober and his colleagues, Fitelson and Stephens, are major opponents of Dembski.[22] Sober subscribes to IBE, in which the best explanation must be selected out of at least two competing hypotheses. Also, he adopts the likelihood approach introduced by Fisher.[23] Unlike the Fisherian significance testing approach, the likelihood approach has no need for significance levels or small probabilities. It is concerned with the probability of the observed data given the hypothesis [P(D|H)], which is not the same as the probability of the hypothesis given the observed data [P(H|D)]. Sober gave this example: Let H be the hypothesis, “There are gremlins in the attic, and they make noise.” It means that if there were actually gremlins in the attic, we would expect to hear noise in that area. In this case, P(D|H) is very high. However, if we hear noise in the attic and guess that the noise is caused by gremlins, this case is P(H|D). This probability is not high at all because the noise could be from something else. [24]

In the significance testing/rejection region approach, the chance hypothesis could be eliminated without accepting another. But in the likelihood approach, the likelihood of a parameter is proportional to the probability of the data, and it gives a function which maximizes the likelihood of observing the data that were actually observed. Fisher called it **the maximum likelihood**. Usually the researcher attempts to find a hypothesis that carries the maximum likelihood, and thus other hypotheses can be eliminated a result of finding the best hypothesis. In other words, D strongly favors $H_1$ over $H_2$ if and only if $H_1$ assigns to D a probability that is much bigger than the probability that $H_2$ assigns to D. In the context of DI, researchers should determine the likelihood of each competing hypothesis (evolution, random process, intelligent design) and determine which hypothesis is the inference to the best explanation. In Sober’s view, DSTA, which is based upon significance testing without consideration of alternatives, departs from the likelihood approach and renders DI untestable and unscientific. The following is a summary of the commonalities and differences between SLA and DSTA:

**Table 1. Commonalities and Differences between SLA and DSTA**

| | **Sober's Likelihood Approach (SLA)** | **Dembski's Significance Testing Approach (DSTA)** |
|---|---|---|
| Sources | Fisher statistics and Inference to the Best Explanation | Fisher statistics |
| Numeric indicator | Likelihood of observing the data that were actually obtained | Probability of the event's occurrence given the chance hypothesis |
| Decision basis | Depends on which hypothesis carries the maximum likelihood among competing hypotheses | Whether the data falls inside or outside the rejection region based upon the cut-off level |
| Conclusion | Favors natural selection | Favors the Design Inference |

Dembski rejects the notion that the best inference must arise from competing explanations by saying that to be forced to choose among hypotheses, on some occasions, is like being forced to choose between "the moon being made entirely of cheese" or "the moon being made entirely of nylon." Dembski further illustrates why choosing competing explanations is unacceptable with the following metaphor:

> Suppose you are the admissions officer at a prestigious medical school. Lots of people want to get into your medical school; indeed, so many that even among qualified applicants you cannot admit them all. You feel bad about these qualified applicants who do not make it and wish them well. Nonetheless, you are committed to getting the best students possible, so among qualified applicants you choose only those at the very top. What's more, because your medical school is so prestigious, you are free to choose only from the top. There is, however, another type of student who applies to your medical school, one whose grades are poor, who shows no intellectual spark, and who lacks the requisite premedical training. These are the unqualified students and you have no compunction about weeding them out immediately, nor do you care what their fate is in

> graduate school. In this analogy the unqualified students are the hypotheses weeded out by Fisher's approach to hypothesis testing whereas the qualified students are those sifted by the likelihood approach. If, perchance, only unqualified students apply one year…the right thing to do would be to reject all of them rather than to admit the best of a bad lot. [25]

It is true that in many situations the so-called competing hypotheses are not qualified to be meaningful explanations. This point has been illustrated in my example of an engineer comparing a Porsche 911 and a Dodge Neon. However, as mentioned earlier, according to Harman one must formulate competing explanations that they are really capable of competition, not just dummy alternative explanations that fulfill the statistical ritual. Nonetheless, the scenario described by Dembski is even worse than dummy alternatives. It is like the aforementioned engineer being forced to accept either a Dodge Neon or a Ford Escort as having superior performance.

What Dembski appeals to is the "bad lot argument," which has been thoroughly discussed by Van Fraassen, Psillos, Lipton, Ladyman et al, and Iranzo.[26] In Dembski's argument, forcing the researcher to accept an inferior explanation out of a bad lot is problematic. However, when we look closely at Peircean abduction, we find that abductive reasoning provides a stepping stone to further pursue a hypothesis with caution, but not to confirm a hypothesis without reservation. As Iranzo states, "When there is comparison and selection, choosing a theory is usually accompanied by believing it, but sometimes it does not happen…being the best explanation does not guarantee justification."[27] In the history of science, sometimes we have to work with weak theories when better options have not yet emerged. For example in the early 20$^{th}$ century, psychometricians adopted a single intelligence factor (g-factor) to conduct factor analysis for the testing of mental capabilities. In later years psychometricians replaced unidimensional factor analysis with multidimensional factor analysis when better theories and methodologies of psychometrics became available. Thus, a tentative acceptance of a weak theory out of a bad lot may pave the way to subsequent advanced theories and methodologies.

Further, even if all competing explanations are low in likelihood due to a bad lot, I see no reason that the researcher must admit a weak explanation, just as the admissions officer does not have to accept unqualified candidates. As a matter of fact, Sober has no problem in rejecting all competing hypotheses: “This principle (likelihood) simply says whether the observations under consideration favor one hypothesis over another. It does not tell you to believe the one that is better supported by the piece of evidence under consideration. In fact, you may, in a given case, decline to believe either hypothesis.” [28]

**Discussion and Recommendations**

Abductive reasoning, advocated by Peirce, is discovery in essence. It is especially applicable to nonrepeatable events, surprising phenomena, and uncertain cases. However, it is doubtful whether applying a Fisherian probability model is appropriate to the issue of DI. When background information is absent, even choosing the proper cut-off level to set up the region of rejection is controversial. Koons suggests that instead of conceptualizing the Dembski/Fisher approach in terms of the empirical/frequentist approach, one could adopt a “rationalist” version of the classical Fisherian model, which would allow the judgment of the chance hypothesis to be based on nonempirical, a priori grounds. This may be a viable solution if and only if the rationalist approach is fully developed.[29]

Unlike the Neyman-Pearson hypothesis testing model, the significance testing approach introduced by Fisher has the null hypothesis only. It is understandable that, following this line of thought, Dembski tends to assert that after eliminating the chance hypothesis there is no need to consider other probability distributions. However, this approach has too much confirmatory character and lacks the Peircean exploratory spirit.

Dembski's argument against SLA seems unconvincing. As Harman emphasizes, IBE seeks out competing explanations rather than just alternative explanations. No serious IBE supporters would consider hypotheses like "the moon is made entirely of cheese" or "the moon is made entirely of nylon." More importantly, Harman defines the goodness of explanations in terms of their fitness to the total explanatory picture of the world. While it is not necessary for the researcher to exhaust all possible competing explanations, at least he/she should relate the explanation to any relevant background information. Dembski, however, tends to de-emphasize the relative and comparative nature of IBE, and puts too much weight on the statistical significance yielded from the rejection region.

Milne correctly pointed out the weaknesses of Dembski's approach:

> What is missing in Dembski's account is any mention of inference to the best explanation. On his account, one tries out the other options, regularity and chance first, and only when these have failed does the possibility of design come into play…Design wins out by default when the other two fail…design isn't really a kind of explanation at all: to fail to explain by regularity or chance is not in itself to succeed with some third kind of explanation. Taking Dembski at his word as to what 'design' really means in his scheme, we can see why there is no mention of inference to the best explanation.[30]

In spite of these shortcomings, Dembski's approach is not beyond repair. As a suggested remedy, the question of Intelligent Design could be re-conceptualized in light of IBE as a supplement to, not as a replacement for, Fisherian significance testing. Abductive reasoning or

IBE is considered the philosophical justification for Exploratory Data Analysis (EDA), in which the primary objective is to examine the data pattern in order to generate a plausible hypothesis.[31] Although at first glance EDA is in opposition to Fisherian hypothesis testing, there is an underlying continuity between the two in terms of their logical reasoning. Josephson and Josephson argued that the whole notion of a controlled experiment, which was also invented by R. A. Fisher, is covertly based on the logic of abduction.[32] In a controlled experiment, the researchers control alternate explanations and test the condition generated from the most plausible hypothesis. However, abduction shares more common ground with EDA than with controlled experiments. In EDA, after observing some surprising facts, the researcher exploits them and checks the predicted values against the observed values and residuals. Although there may be more than one convincing pattern, we "abduct" only those that are more plausible for subsequent inquiry.

Some may wonder how abductive reasoning is different from Dembski's approach, since in his scheme Intelligent Design emerges to be a more plausible hypothesis after regularity and chance are filtered. It is important to point out that in abductive reasoning, IBE and EDA **go from data to hypotheses,** while inductive reasoning based upon the conventional Fisherian approach goes from hypothesis to expected data. It is recommended that the issue of Intelligent Design should be addressed in the abductive mode, which places the emphasis on hypothesis generation instead of elimination based upon a pre-determined distribution and rejection region.

Last but not least, unlike what Dembski suggests, the Fisherian and Bayesian approaches are not necessarily mutually exclusive. For example, Pawitan also attempted to synthesize the frequentist and the Bayesian approaches. Although Pawitan also viewed probability as a measure of uncertainty, he accepted a "ladder of uncertainty," a Fisherian idea introduced in his last book

*Statistical methods and scientific inference*: Whenever possible, the researcher should base inference on probability statements; otherwise, it should be based on likelihood.[33] With the ladder of uncertainty as the foundation, Pawitan proposed the likelihood approach: Uncertainty can be expressed by both likelihood and probability, where likelihood is a weaker measure of uncertainty and probability allows objective verification in terms of long term frequencies. Pawitan argued that the likelihood approach is a compromise between Bayesianism and frequentism because this approach carries features from both factions.

Further, Pawitan developed the empirical likelihood approach by merging the likelihood and the bootstrap, which is a resampling method. In bootstrap, the sample is duplicated many times and treated as a virtual population. Then samples are drawn from this virtual population to construct an empirical sampling distribution. Pawitan is not the only one who synthesizes the Fisherian and the Bayesian school, but this is one of many examples to demonstrate the notion that making the Fisherian and the Bayesian as an “either-or” question is out-dated. Rather, the question is one of giving a Bayesian account of explanatory virtues, which is now perhaps the most vital research program in the epistemology of scientific inference. It is my belief that taking the Bayesian approach into account, as well as forming the hypothesis by IBE, can advance the development of the Intelligent Design Argument.

**Acknowledgements**

Special thanks to Dr. Brad Armendt for his valuable input and Ms. Samantha Waselus for her professional editing.

**End notes**:

---

[1] Examples of opponents are Elliot Sober, " Testability." *Proceedings and Addresses of the American Philosophical Association* 73 (1999): 47-76; Elliot Sober, *Philosophy of Biology* (2nd ed.) (Boulder, CO: West View Press, 2000); John Mackie, *The Miracle of Theism*. (Oxford: Oxford University Press, 1982). Examples of proponents are Stephens Meyer, "The Scientific Status of Intelligent Design," in *Science and Evidence for Design in the Universe*, edited by W. Dembski & S. Meyer, (San Francisco, CA: Ignatius Press, 1999), 151-212; Richard Swinburne, *The Existence of God*. (Oxford: Oxford University Press, 1991). For abductive reasoning, please consult Charles Sanders Peirce, *Collected papers of Charles Sanders Peirce*. (Cambridge: Harvard University Press, 1934/1960). For IBE, please consult Gilbert Harman, "The Inference to the Best Explanation." *Philosophical Review* 74 (1965): 88-95; Gilbert Harman, "Enumerative Induction as Inference to the Best Explanation," *Journal of Philosophy* 65 (1968): 529-533.

[2] William Dembski, *The Design Inference: Eliminating Chance through Small Probabilities*. (Cambridge: Cambridge University Press, 1998); *Intelligent Design: The Bridge between Science and Theology*. (Downers Grove, IL: InterVarsity Press, 1999); *No Free Lunch: Why Specified Complexity cannot be Purchased without Intelligence*. (Oxford: Rowman & Littlefield, 2002); *The Design Revolution: Answering the Toughest Questions about Intelligent Design*. (Downers Grove, IL: InterVarsity Press, 2004).

[3] Robert Koons, "Are Probabilities Indispensable to the Design Inference?" *International Society for Complexity, Information, and Design*. [database online] : http://www.utexas.edu/cola/depts/philosophy/faculty/koons/ontocomplex.pdf (Accessed on July 30 2001).

[4] Chong Ho Yu, "Testability of the Intelligent Design Argument in the Perspective of Quantitative Methodology," *Jian Dao Journal* 20 (2003): 89-102.

[5] Guy Deutscher, "On the Misuse of the Notion of 'Abduction' in Linguistics," *Journal of Linguistics* 38(2002): 469-485.

[6] William Rozeboom, , "Good Science is Abductive, not Hypothetico-Deductive." In *What if There were no Significance Tests?* edited by Lisa L. Harlow, Stanley A. Mulaik, & James, H. Steiger, (Mahwah, NJ: LEA Publishers, 1997), 335-392, p.368.

[7] John Behrens, & Chong-Ho Yu, "Exploratory Data Analysis." In *Handbook of Psychology Volume 2: Research Methods in Psychology*. Edited by J. A. Schinka & W. F. Velicer, (New Jersey: John Wiley & Sons, Inc., 2003), 33-64, p.41.

[8] Chong-Ho Yu, *Philosophical Foundations of Quantitative Research Methodology*. (Lanham, MD: University Press of America, Forthcoming).

[9] Hintikka, 1998, "What is Abduction? The Fundamental Problem of Contemporary Epistemology." *Transactions of the Charles S. Peirce Society* 34(1998): 503-533, p.524

[10] Jaakko Hintikka, 1998, 523.

[11] Harman, 1968, 530-533.

[12] Chong-Ho Yu, Experimental Design as Variance Control. [Database online] http://www.creative-wisdom.com//teaching/WBI/variance_control.html (Accessed on July 20, 2005).

[13] Jonathan Treitel, "Confirmation as Competition: The Necessity for Dummy Rival Hypotheses." *Studies in the History and Philosophy of Science* 18(1987): 517-525.

[14] Dembski, 2004, 232-233.

[15] For evidence that Bayesian inferences are widely accepted by the statistical community, please consult James Berger, "Bayesian Analysis: A Look at Today and Thoughts of Tomorrow." *Journal of American Statistical Association* 95(2000): 1269-1276; Yudi Pawitan., *In all likelihood: Statistical Modeling and Inference using Likelihood*. (New York: Oxford University Press, 2001). Coles and Lucchin applied Bayesian methods into the study of the origin of the universe. Please consult Peter Coles & Francesco Lucchin, *Cosmology: The Origin and Evolution of Cosmic Structure*. (New York: Johnson Wiley & Sons, 2002).

[16] James Press & Juith Tanur, *The Subjectivity of Scientists and the Bayesian Approach*. (New York: John Wiley & Sons, Inc, 2001), 204.

[17] Chong-Ho Yu, "Resampling Methods: Concepts, Applications, and Justification." *Practical Assessment Research and Evaluation*, 8(2003), [Database online]: http://pareonline.net/getvn.asp?v=8&n=19

[18] Bruce Thompson, "Statistical, Practical, and Clinical: How Many Kinds of Significance do Counselors need to Consider?" *Journal of Counseling and Development* 80(2002), 64-71.

[19] E. L. Lehmann, "The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two?" *Journal of the American Statistical Association* 88 (1993), 1242-1249; Chong-Ho Yu, "History of Science and Statistical Education: Examples from Fisherian and Pearsonian Schools," *2004 Proceedings of the American Statistical Association, Statistical Education Section* [CD-ROM], Alexandria, VA: American Statistical Association.
[20] Dembski, 1998, 41-42.
[21] Ralph Rosnow & Robert Rosenthal, "Statistical Procedures and the Justification of Knowledge in Psychological Science." *American Psychologist* 44 (1989): 1276-1284, p.1277
[22] Branden Fitelson , Christopher Stephens , & Elliot Sober, "How not to Detect Design – Critical Notice: William A. Dembski, The Design Inference," *Philosophy of Science* 66(1999): 472-488.
[23] Ronald. A. Fisher, Ronald, *Contributions to Mathematical Statistics*. (New York: Wiley, 1921/1950).
[24] Sober, 2000, 32-33.
[25] Dembski, 2002, 104-105.
[26] Bas Van Fraassen, *Laws and Symmetry*. (Oxford: Oxford University Press, 1989); Stathis Psillos, "On Van Fraassen's Critique of Abductive Reasoning," *Philosophical Quarterly* 46(1996): 31-47. Peter Lipton, "Is the Best Good Enough?" In The Philosophy of Science, edited by D. Papineau, (Oxford: Oxford University Press, 1996), 93-106. James Ladyman, Igor Douven, Leon Horsten, & Bas Van Fraassen, "A Defense of Van Fraassen's Critique of Abductive Inference: Reply to Psillos." *Philosophical Quarterly* 47(1996): 305-321. Valeriano Iranzo, "Bad Lots, Good Explanations." *Critica* 33(2001): 71-96.
[27] Iranzo, 2001, 88-89.
[28] Sober, 2000, 32.
[29] Koons, 2002.
[30] Peter Milne, "Review: William A. Dembski's The Design Inference," *British Journal for the Philosophy of Science* 52(2001): 801-808, p.805.
[31] Chong-Ho Yu, "Induction? Deduction? Abduction? Is there a Logic of EDA?" Paper presented at the Annual Meeting of American Educational Research Association, (New Orleans, LA, 1994, April). (ERIC Document Reproduction Service No. ED 376 173).
[32] John Josephson & Susan Josephson, *Abductive Inference: Computation, Philosophy, Technology*. (Cambridge, UK: Cambridge University Press, 1994).
[33] Ronald A. Fisher, *Statistical Methods and Scientific Inference*. (New York: Hafner, 1956).