RUNNING HEAD: ID Testability

Testability of Intelligent Design Argument in the Perspective of Quantitative Methodology

Chong Ho Yu, Ph.D., CCNA, MCSE, CNE

Printed in Jian Dao Journal, 20 (2003): 89-102

Chong Ho Yu, Ph.D.
Psychometrician
PO Box 612
Tempe AZ 85280
Email1: asumain@yahoo.com.hk
Email2: chonghoyu@yahoo.com
Website: http://seamonkey.ed.asu.edu/~alex/pub/pub.html

Abstract

Renowned philosopher of biology Elliott Sober argued that some hypotheses defended by creationists are testable, yet auxiliary assumptions associated with the intelligent design argument cannot be evaluated. Sober stated that judging by the current status and the pace of progress of creationism, it is doubtful whether creationism can be treated as a serious scientific research program. Further, based on Fisher's likelihood principle, Sober argued against the probabilistic inference suggested by the design hypothesis. In this paper it is argued that the testing procedure and the probabilistic inference for the intelligent design hypothesis conform to the standards of quantitative research methodology. Moreover, if the evaluation of research progress is centered on how well the design argument is articulated, creationism indeed is dynamic and progressive. Other arguments given by Sober will also be examined.

Testability of Intelligent Design Argument in the Perspective of Quantitative Methodology

Chong Ho Yu, Ph.D.

Although many scholars (e.g. Meyer, Dembski, Behe) have argued that the intelligent design (ID) hypothesis (creationism) is just as good a science as any other, and that not to consider creationism would be to present evolution dogmatically, many researchers (e.g. Sober, Kitcher) have raised questions on its testability in a scientific sense. This article concentrates on arguments introduced by Sober and points out several shortcomings of his argument.

Creationism is usually perceived by scientists as being affiliated with "Christian fundamentalism" and thus its inquiry is not seriously considered.[1] For instance, prominent geologist Newell simply denounced creationism as non-science:

> Creationism cannot be defined as a science because it is based on an inflexible presupposition, a conviction based on supposed cases and events that cannot be examined either directly or indirectly by scientific methods. The conclusion precedes the search for evidence. In addition, there is no scientific research into the basic tenets of creationism. Revelation is completely outside the scope of science. The 'research' of creationism is a biased, destructive critique of all scientific investigation of origins. The basic position of creationism is negative, that is, creationists seek to enhance their credibility by destroying the opponent, not by building their own case. [2]

Some scholars have viewed fundamentalist creationism as science, but just a bad science.[3] They argue that in the face of unanswered questions good scientists look for more data, but creationists rest their answers on the supernatural. However, when people argued against creationism, they might not have the same definition of creationism in their mind. Indeed, there are many other versions of creationism. Some versions believe in geological findings instead of insisting upon a "young earth" argument. Some versions accept micro-evolution within species

instead of embracing the notion of fixed categories of species. In this paper, instead of addressing the merits and weaknesses of every version of creationism, a global definition of creationism is used. This global definition identifies creationism as the affirmation that the universe and species originated from an intelligent designer. Thus, in this article the terms "creationism" and "intelligent design argument" are used interchangeably. The objective of this paper is to discuss whether this intelligent design argument qualifies as scientific inquiry.

In order to answer the preceding question, one must attempt to operationally define science. Conventionally, science is defined as a systematic inquiry process that involves empirical studies of a phenomenon or group of phenomena for testing hypotheses from predictions based upon generalized laws.[4] Needless to say, this definition is open to debate. Unless Monte Carlo simulations are regarded as empirical methods, mathematics could not be credited as science by this definition because mathematics does not utilize empirical proof. Psychology, especially cognitive psychology, uses latent constructs rather than observed items as variables; if the scope of science is confined in observed phenomena, psychology is disqualified as a scientific discipline, too.

Indeed, both evolutionary biology and ID face difficulties in meeting the standards listed by the preceding definition of science. The requirement of predictions is a major challenge. According to Hempel, a scientific explanation is a "deductive-nomological" (DN) process that involves: (1) initial conditions, and (2) law-like generalizations. One can predict that a phenomenon will occur if the initial conditions are right based on the universal laws.[5] Evolution is criticized for not being a predictive science and for not having general laws. Wilkins pointed out that unlike physicists, evolutionists cannot predict with any reasonable degree of accuracy what mutations will arise, which genotypes will recombine, and what other events will affect the way species develop over time.[6] Moreover, the so-called '"laws" of genetics and other biological rules are not laws in a strict sense. Further, as a tool of evolutionary research, the primary mission of phylogeny is to reconstruct the past rather than to predict the future. In fact, creationism also

suffers from the same kind of shortcomings. Theologians could not tell what the initial conditions were like in the beginning of the universe. Also, no one could conceive whether God is governed by any laws in the process of creation. And like phylogeny, creationism is retrospective instead of predictive.

In this paper, science is defined as an inquiry process that involves both logical reasoning and empirical methods of observed phenomena and theoretical entities for hypothesis testing. This is not a hasty definition to relax the game rules for both evolutionary biology and creationism. Rather, based upon the unsuitability of the traditional definition of science to many disciplines, a new definition is needed. Indeed, many modern good sciences do not rest on deterministic laws, but on propensities, likelihood, and probability. Whether creationism should be taken as a serious scientific inquiry depends on how well the following questions are answered: (1) Are the hypotheses suggested by creationism testable? (2) Is the logical reasoning applied to the design argument as valid as that of other sciences? (3) Could probabilistic inferences, which are employed by other sciences, be well-applied to creationism?

Renowned philosopher of biology Elliott Sober argued that some hypotheses defended by creationists are testable, yet auxiliary assumptions associated with the intelligent design argument cannot be evaluated.[7] Sober stated that, based on the current status and the pace of progress of creationism, it is doubtful whether creationism can be treated as a serious scientific research program. The arguments given by Sober will be examined one by one.

Testability and falsifiability

Sober rejected the principle of falsifiability introduced by Popper as a criterion to judge whether a proposition is scientifically testable.[8] To Popper, scientific propositions can be falsified empirically. On the other hand, unscientific claims are always right and cannot be falsified at all. In this view, Freudian psychoanalysis, Adlerian psychology, Marxism, creationism, and evolutionary biology do not belong to the domain of science because they are capable of fitting theories with many possible observations. Sober pointed out several problems with the

Popperian notion. One problem is that the principle of falsifiability would disallow probabilistic statements. For instance, assuming that a coin is fair, what is the probability of it landing heads when tossed ten times? It is possible for a fair coin to land heads on all ten tosses, or to land heads on nine and tails on one. In other words, the hypothesis that the coin is fair is compatible with all possible outcomes. Therefore, probabilistic statements are not falsifiable in Popper's sense.

It is difficult, if not impossible, to test or falsify evolution and creationism empirically. Weiner admitted, "Darwin himself never tried to produce experimental confirmation of this particular point [evolution]. It is at once extremely logical and extremely hard work to prove."[9] Empirical studies such as those conducted by the Grants demonstrated the adaptation of species. But the notion of evolution, in which new species could evolve from other species, is never empirically substantiated. In the Popperian view, adaptation is hardly evidence of evolution. Bair was critical of evolution because all observations could fit into the evolutionary framework, "If I observe a species change, that proves evolution. If I see a species perservere, that is natural selection which also proves evolution."[10] In a similar vein, Sober pointed out that creationism is always right if the answers "God did it" or "It's God's will" are provided.[11] Again, all observations can fit into the theological framework. In short, if the Popperian view is adopted as a standard of testability, both creationism and evolution should not be accepted as scientific inquiries. Sober made a smart move to evaluate scientific propositions in terms of probability. This point will be discussed in a later section.

Duhem thesis and auxiliary assumptions

Sober further criticized Popper by using the Duhem thesis.[12] Popper held that there is an asymmetry between falsification and verification. It is possible to conclusively prove theories false but impossible to conclusively prove them true. For example, if test scores do not improve after a treatment is introduced, the treatment can be judged as ineffective. But if the test scores improve, one cannot assert that the treatment must be working while many relevant conditions

that may affect the treatment outcome are not examined. Popper's criticism of verification is legitimate. French physicist and philosopher Duhem suggested that a hypothesis should not be tested individually. Rather, a web of theories, including the hypothesis and auxiliary assumptions, should be tested in a holistic manner.[13] Sober argued that by inserting auxiliary assumptions into testing, the symmetry between falsification and verification could be restored.

Sober made a revision of Duhem's holistic view by asserting that auxiliary assumptions are often independently tested in science.[14] For example, when a woman goes to a clinic for a pregnancy test, the auxiliary assumption, that the test procedure is reliable, has been verified before the woman is tested for pregnancy. Sober complained that there is no way to evaluate the likelihood of the auxiliary assumptions associated with the creation hypothesis. To hypothesize that God created the universe and species, one must also specify the probability that species with such and such features would be created based on God's ability and his intention. However, different religions conceive of God in different ways. How can one know which auxiliary assumption about God is believable? Because Sober contended that it is an important part of scientific practice that conjunctions can be broken apart, he doubted the scientific value of creationism.[15] Again, Sober used a clinical test as a metaphor:

> If your doctor runs a test to see whether you have tuberculosis, you will not be satisfied if she reports that the conjunction "you have tuberculosis and auxiliary assumption 1" is very likely while the conjunction "you have tuberculosis and auxiliary assumption 2" is very unlikely. You want your doctor to address the first conjunct, not just various conjunctions. And you want her to do this by using a test procedure that is independently known to have small error probabilities. Demand no less of your theologian.[16]

This argument is problematic. First, variables cannot be as easily pulled apart as Sober described. Many scientific research findings are conjunctive in nature. In both clinical examples, Sober equated auxiliary assumptions to test reliability, which can be assessed independently. However, when auxiliary assumptions are concerned with variables and covariates, quantitative

researchers usually employ multivariate procedures because the main hypothesis and its auxiliary assumptions cannot be easily separated, and their interaction effects must be assessed in a holistic fashion. Take multiple regression as an example. When five regressors are entered into the model and only three regressors are found as significant predictors, a naive researcher may throw away two other variables and report the three significant ones. Indeed, the rationale of using multiple regression is to test all predictors simultaneously, and thus the result should be interpreted as a whole model. If another regression is run with the three significant variables only, their p values and slopes may change substantively.

Second, it is surprising that Sober argued for a simple (univariate, dichotomous, or both) answer while he endorsed probabilistic inferences elsewhere. The TB and pregnancy tests mentioned by Sober are metaphors. Whether it is a sound argument depends on how strong the metaphorical link is. However, scientific research is by no means similar to a physical check-up. In a physical check-up a clear-cut answer is given to a single patient. In scientific research many cases are evaluated in order to draw a probabilistic inference to a theory. When there are many cases, variances and residuals are expected and thus the conclusion is hardly clear-cut.

Third, even if we use the test procedure reliability as one of the auxiliary assumptions, this auxiliary assumption carries other auxiliary assumptions that cannot be tested independently. For example, Cronbach Coefficient Alpha is one of the measures that estimate reliability in terms of internal consistency. Cronbach Alpha is based upon the true score model, in which assumptions cannot be tested. The theory behind Cronbach Alpha is that the observed score is equal to the true score plus the measurement error ($Y = T + E$). For example, I might know 80% of the material but get a score of 85% because of guessing. In this case, my observed score is 85 while my true score is 80. The additional five points are due to the measurement error. It is assumed that a reliable test should minimize the measurement error so that the error is not highly correlated with the true score.[17] There are two assumptions here. First, it is assumed that the true score really exists. But what is "true" remains an unproven ontological assumption.

Second, it is assumed that the same examinee would show a within-subject variability if he/she retook the same test over and over. This point is illustrated by a thought experiment. Assume that a person takes a test. Afterwards, we wipe out his/her memory about the test questions and let him/her retake the same test. Even though the examinee retains the same level of ability, his/her score on the second test would not be the same as that on the first one. However, since no one is able to erase human memory, the auxiliary assumption of the test procedure reliability is not empirically tested (perhaps we can erase all memories of a person, but it is difficult to erase one's memory about the test only while all other skills are retained) .

Fourth, auxiliary assumptions are like hypotheses. There are always competing conceptions of God. But is it true that there are also competing auxiliary assumptions relating to evolution? Using the test procedure reliability as an example, it is doubtful whether the scientific community has reached a common consent on dating methods. It is also questionable if the test of the evolutionary hypothesis can be boldly claimed as a test that is independent of auxiliary assumptions. In short, imposing the criterion of testing auxiliary assumptions independently on creationism is unfair.

Likelihood principle and probabilistic inferences

As mentioned before, Sober rejected the Popperian notion for its limitation on evaluating probabilistic inferences. Instead, Sober applied Fisher's likelihood principle, which is an approach of probabilistic inference, to question the validity of creationism.[18] According to the likelihood principle, the probability of the observed data given the hypothesis [P(D|H)] is not the same as the probability of the hypothesis given the observed data [P(H|D)]. The former is about likelihood, while the latter is concerned with probability. For example, Let H be the hypothesis, "There are gremlins in the attic, and they make noise." It means that if there actually were gremlins in the attic, we would expect to hear noise. In this case, P(D|H) is very high. However, if we hear noise in the attic and guess that the noise is from gremlins, this case is P(H|D). This probability is not high at all because the noise could be from something else.

Using this likelihood principle, Sober criticized that the logic of the design argument rests on P(H|D) rather than P(D|H). If we know that an intelligent designer would make the universe in such and such a way, then we could argue that since the world appears in this order, it is likely that this world was created by an intelligent designer. In this case, the probability is P(D|H). However, P(H|D), instead of P(D|H), is the underlying logic of the design argument. The design argument starts from the existing world, which appears to be well-structured. Given the existing world, creationists argue that the world originates from an intelligent designer. According to Sober, this problem is like the earlier example that when we hear noise in the attic, we assert that the noise is made by gremlins.

At first glance, Sober's argument is convincing. As a matter of fact, computing P(D|H) is a standard practice among quantitative researchers. Many statistic procedures aim to find out the probability that the observed data would occur if the null hypothesis were true. However, this approach has been attacked by many statisticians and social scientists.[19] Interestingly enough, Carver made the following comments about P(D|H):

> What is the probability of obtaining a dead person (D) given that the person was hanged (H); that is, in symbol form, what is P(D|H)? Obviously, it will be very high, perhaps .97 or higher. Now, let us reverse the question: What is the probability that a person has been hanged (H) given that the person is dead (D); that is, what is P(H|D)? This time the probability will undoubtedly be very low, perhaps .01 or lower. No one would be likely to make the mistake of substituting the first estimate (.97) for the second (.01); that is, to accept .97 as the probability that a person has been hanged given that the person is dead.[20]

Was Carver making a negative comment against ID? No, this paragraph is extracted from Carver's article entitled "The Case against Statistical Testing." It is a criticism against the Fisherian hypothesis testing employed by many scientists!

P(D|H) could be viewed as hypothesis-driven, while P(H|D) could be regarded as data-driven. Some critics have said that the primary question in many research studies should be the other way around: Given the observed data, what is the probability that the hypothesis is true? In recent years several data-driven approaches such as Exploratory Data Analysis (EDA) have been developed as supplements to the conventional hypothesis testing. If the design argument is said to lack scientific merit due to the use of P(H|D), then several other data-driven methodologies are also questionable. In short, like many other scientific inquiries, the ID hypothesis should not be devalued by employing P(H|D).

Indeed, in many scientific inquiries a well-formulated hypothesis and prior knowledge are absent. In this case, it is legitimate to assign a subjective probability for estimation. By the same token, the intelligent design argument could be reasoned with the application of the Bayesian inference.[21]

Moreover, Sober criticized that creationists unfairly compared natural selection with the mindless random hypothesis.[22] Sober argued that natural selection is not a random process. If a process is random, then different options have the same probabilities. However, the principle of natural selection implies that the fittest species have the highest probability of survival. Thus, natural selection is not a chance process, and the comparison made by creationists is invalid.

It is puzzling that on one hand P(D|H) is highly regarded by Sober, but on the other hand he objected to a comparison based upon a random process. The testing of P(D|H) is, given that the null hypothesis is true, the likelihood that the observed data would be obtained by random chance alone. In this context, chance fluctuation is synonymous with random fluctuation. Many educational and psychological studies use an inferior treatment or a placebo treatment for the control group. Comparing the performance of the treatment group with that of the control group tends to yield a significant treatment effect.[23] Is comparing the design hypothesis against the random process hypothesis the same type of mistake as that committed by those educational researchers and psychologists? Whether natural selection is a random process is one question.

Whether comparing the design hypothesis with a random process is fair is another question. If a creationist does not explicitly declare that natural selection is a random process, the comparison based upon random fluctuations is entirely valid. In some psychological studies, there is no control group and all rival treatments are equally well-developed. In spite of this, the testing of P(D|H) still aims to find out the probability that the observed difference just happens by chance alone. In short, the following question raised by creationists is scientifically legitimate: "Given the observed structure and function of species, what is the likelihood that it occurs by a random process?"

Research tradition and scientific progress

Laudan stated that besides the problem-solving ability, the rate of progress of a research tradition should be taken into consideration when assessing the scientific merits of a research tradition.[24] Based on Laudan's notion, Sober charged that creationism is not a promising scientific research tradition:

> Present-day evolutionary theory has formulated and tested countless hypotheses of which Darwin never dreamed. Present-day creationism, however, is much like old-time creationism in that the basic claim God created this and that feature of the living world has not been elaborated and extended. Genuinely scientific theories are extended and refined over time in ways that allow new observations to be brought to bear. The intellectual stagnation that one finds in creationist thought is a sign that something has gone wrong…It is important to realize that creationism is effective not only in its current theories but in its historical track record: It is no surprise that biologists have come to regard 'creation science' as a contradiction in terms.[25]

It is true that the evolutionary theory has changed a lot since Darwin. However, it is doubtful whether all these changes can be counted as "progress." For example, the interpretation of Pithecanthropus, which was discovered by Dr. Dubois in Java, had been changed many times.

At the end Dr. Dubois retracted the evidence of Pithecanthropus. No wonder Heinze mocked that the only thing that has evolved is the theory of evolution. [26]

If the objective of the inquiry concerning the origin of species is to give dates and details, there is no doubt that creationism is a stalled research tradition. The Bible does not give instructions about Carbon 14 dating and DNA sequencing. However, do the problems of the dates and details of the origin of species belong to the realm of creationism? Asking theologians to answer geological and biological questions is like demanding painters to explain how the chemical reaction in the mixture of oil paints and oil thinner works, asking photographers to explain how the electrical circuits in a CCD sensor function in a digital camera, or asking statisticians to describe how the Intel Itanium processor handles random number generation. The existence of an intelligent designer is central to the Christian doctrines, and thus it is understandable for Christians to defend the design hypothesis. For the Christian church, progress is evaluated by how well this position is defended. Indeed, the arguments suggested by Anselm, St. Thomas Aquinas, Pascal, Descartes, Kant, Barth, and Tillich are very different. Labeling the design hypothesis as a non-progressive research tradition is unfair. Moreover, creationism is only part of a rich heritage of Christian theology. Throughout two thousand years theology has made a lot of progress and has contributed to human inquiries in different ways.

Conclusion

Creationism, which argues for the existence of an intelligent designer, is a testable hypothesis. Although the design hypothesis is not testable under the Popperian notion, it is testable by probabilistic inferences. By applying the Duhem thesis, Sober questioned the testability of creationism because auxiliary assumptions related to the design hypothesis cannot be examined independently. Further, based on Fisher's likelihood principle, Sober argued against the probabilistic inference suggested by the ID hypothesis. However, the testing procedure and the probabilistic inference for the design hypothesis conform to the standards of quantitative research methodology. Sober also charged that creationism is not progressive, but it

depends on how progress is evaluated. If the evaluation of progress is centered on how well the design argument is articulated, creationism indeed is dynamic and progressive.

Creationists and evolutionary biologists should not attempt to mute each other. When both views are presented to the audience, insightful findings may result from the interaction between these two theories. For instance, in the early part of last century the Catholic theologian Teilhard de Chardin attempted to synthesize creationism and evolution.[27] However, the Catholic Church issued a monitum (a red flag caution) against acceptance of Charin's thesis. Later the Catholic Church lifted the monitum and praised his work as an inspiration of progressive theology. In the history of science quite a few valuable research traditions are a fusion of incompatible theories. For example, in early last century the Fisherian model and the Neyman/Pearson model had many incompatible elements. Fisher and Pearson didn't even talk to each other because of their uncompromising insistence on their own perspectives. Nonetheless, the current hypothesis testing model is a synthesis of the works of Fisher, Neyman, and Pearson.[28] If Fisher and Pearson had kept an open dialog and even worked together, the history of statistics would have been very different. Today we need options, openness, and triangulation.

End Notes

1. Vawter, B. Creationism: Creative misuses of the Bible, Is God a creationist? The religious case against creation-science ed. R. M. Frye (New York: Charles's Sons, 1983), 71-82.
2. Newell, N. D. Creation and evolution: Myth or reality? (New York: Columbia University Press, 1982), 16.
3. Shanks, N., & Jolin, K. H., Redundant complexity: A critical analysis of intelligent design in biochemistry. Philosophy of Science 66 (1999): 268-282; Palevitz, B. A., Falling off a tightrope: Compromise and accommodation in the war between creationism and evolution. Bioscience 50 (2000): 926-930; Pigliucci, N., Chance, necessity, and the war against science. Bioscience 50 (2000): 79-81.
4. J. P. Siepmann, What is science? Journal of Theoretics vol. 1, no. 3 (1999), [database on-line] Available from: http://www.journaloftheoretics.com/Editorials/Editorial%201-3.html (accessed 29 April 2001).
5. Hempel, C. G., Aspects of scientific explanation and other essays in the philosophy of science (New York: Free Press, 1965), 331-496.
6. Wilkins, J. Evolution and philosophy [database on-line] Available from: http://www.talkorigins.org/faqs/evolphil/predict.html (accessed 29 April 2001).
7. Sober, E., Testability. Proceedings and Addresses of the American Philosophical Association 73 (1999): 47-96; Sober, E., Philosophy of biology (2nd ed.) (Boulder, CO: West View Press, 2000), 46-57; Sober, E., The design argument. The Blackwell guide to philosophy of religion ed. W. Mann (in press), 1-37.
8. Sober, Philosophy of Biology, 46-52.
9. Weiner, J., The beak of finch: A story of evolution in our time. (New York: Vintage Books, 1995), 52.
10. .Bair, J. (1998). The bogus logic of the Beak [database on-line] Available from: http://ourworld.compuserve.com/homepages/jbair/beakfinc.htm (accessed 29 April 2001).

11. Sober, Philosophy of Biology, 56; Sober, The Design Argument, 7.

12. Sober, Philosophy of Biology, 49.

13. Duhem, P. M. M., The aim and structure of physical theory (Princeton: Princeton University Press, 1954), 183-187.

14. Sober, Testability, 55-57.

15. Sober, Testability, 55-57; Sober, Philosophy of Biology, 49-52.

16. Sober, The Design Argument, 14..

17. Yu, C. H., An Introduction to computing and interpreting Cronbach Coefficient Alpha in SAS. Proceedings of 26th SAS User Group International Conference, 2001, paper 246.

18. Sober, Philosophy of Biology, 14-18.

19. Berger, J. O., & Berry, D. A., Statistical analysis and the illusion of objectivity. American scientist, 76 (1988):159-165; Cohen, J., The earth is round (P < .05). American Psychologist 45 (1994): 1304-1312; Loftus, G., Psychology will be a much better science when we change the way we analyze data. Current Directions in Psychological Science 5 (1996): 161-170.

20. Carver, R.P., The case against statistical testing. Harvard Educational Review 48 (1978): 378-399.

21. Swinburne, R., The existence of God (Oxford: Clarendon Press, 1991), 102-107.

22. Sober, Philosophy of Biology, 36-39.

23. Yu, C. H., Experimental design as variance control [database on-line] Available from: http://seamonkey.ed.asu.edu/alex/teaching/WBI/variance_control.html (accessed 9 August 1999).

24. Laudan, L., Progress and its problems: Towards a theory of scientific growth (Berkeley, CA: University of California Press, 1977), 121-151.

25. Sober, Philosophy of Biology, 27, 57.

26. Heinze, T. Present day example of evolution [database on-line] Available from: http://www.creationism.org/heinze/b19_moth.htm (accessed 29 April 2001).

27. Teilhard de Chardin, P., The appearance of man (New York: Harper & Row Publisher, 1965), 271-273; Teilhard de Chardin, P., Christianity and evolution (New York: Harcourt Brace Jovanovich, 1971), 76-95.

28. Lehmann, E. L., The Fisher, Neyman-Pearson theories of testing hypotheses: One theory or two? Journal of the American Statistical Association 88 (1993): 1242-1249.