Troy D. Riggs, Professor of Mathematics, Union University

Review of Andrew M. Hartley's *Christian and Humanist Foundations for Statistical Inference: Religious Control of Statistical Paradigms*

In this work, Andrew Hartley addresses the well-known debate between the frequentist and the Bayesian interpretations of statistical inference. However, his intent, as expressed in the preface, is "to examine ways religious beliefs control statistical theorizing." His conviction is "that having the right God is a necessary foundation for every scientific endeavor" (p. viii). This core idea is behind the *Philosophy of Law Idea* (PLI) in general, and the author's application of the PLI to statistical inference, in particular.

Now, I grew up around farmers and loggers who when confronted with a mechanical device that wasn't working would say things like "I'm thinking that the problem here has to be with…." I gathered from these discussions that while these men knew that there was an actual problem (reality), reasonable men recognize that judgments are conditional (based on knowledge, experience and individual thought process) and are thus provisional. I was then surprised to learn from my university education that both theoretical and practical approaches to epistemology in science and philosophy don't often show the same discretion.

The PLI, developed by Dutch Calvinist philosophers Herman Dooyeweerd and Dirk Vollenhoven, is a philosophy of science that explicitly acknowledges that God is sovereign over all of creation and that there are multiple aspects of revealed reality (quantitative, spatial, kinematic, etc.) without any one dominant over the others. The PLI takes concrete steps to temper the process of scientific and statistical investigation with explicit humility. While this seems laudable to me, claims about the necessity of having the right God for every scientific endeavor seem a bit excessive. I suspect that many of the laborers whom I watched discussing the potential causes of failed machinery had conceptions of God that would disturb a Dutch Calvinist theologian. However, they usually managed to avoid placing too much confidence in any one line of reasoning.

Having acknowledged this reservation, let me hasten to say that Hartley thoughtfully explores what PLI's principle of multiple aspects of revelation has to say about statistical inference. He defines the term *statistical inference* several times with minor variations, but essentially its purpose is to "provide fiduciary meaning about logical hypotheses, given quantitative data" (p. 82). Or more specifically, that given a set of observed quantitative data, statistical inference should provide a reasonable assessment (measurement) of the credibility (or plausibility) of certain hypotheses.

It seems to me that there is both a weak formulation and a strong formulation of this statement. In the weak formulation of statistical inference, we leave out the word measurement. This is the formulation that I attempt to teach to my introductory statistics students. I emphasize that a (Neyman-Pearson) hypothesis test is a decision-making process where we assume one belief without proof (the null hypothesis, H_0) and put the burden of proof entirely on another competing belief (the alternative hypothesis). The process of attempting to falsify H_0 with evidence results in a decision. The only meaning that we assign to the so-called p-value is that it represents the probability that a true H_0 would produce a result at least as different from expected results as the data x at hand, i.e. $Pr(x|H_0)$.

However this interpretation does not provide a direct measurement of the plausibility of each of the competing hypotheses. Students and practitioners alike would like to read more meaning into the results of a hypothesis test in general and into p-values in particular. We would prefer to make a statement about the probability of the null hypothesis, given the data that we have observed, i.e. find $Pr(H_0|x)$. However, the p-value represents the reverse conditioning, namely $Pr(x|H_0)$, as described above.

At worst, practitioners confuse the two—the "fallacy of the transposed conditional." However, at its best, $Pr(x|H_0)$ is interpreted as an indirect 'likelihood' indicator that can reasonably increase or decrease our confidence in H_0. How this is handled varies widely, however. In Bayesian approaches, prior probabilities and likelihoods are explicitly combined via Bayes' rule. Hartley describes how frequentist approaches may still utilize this information, but through subjective judgments external to the test itself, rather than an explicit incorporation into a measurement of belief.

A quick example, provided by Hartley, is in order. Suppose that a patient presents himself to a physician with spots on his skin. The physician wants to determine whether the patient has measles. While Pr(spots|measles) is fairly high, the Pr(measles|spots) may not be very large, since there are many conditions that produce the symptom of spots on the skin. However, if a physician suspects measles independently of this symptom, the symptom may reasonably push us toward the decision that the patient has measles. This can be done by combining the prior belief in measles explicitly using Bayes rule. It can also be done by comparing the relative size of Pr(spots|measles) to Pr(spots|not measles)—the so-called likelihood ratio. Or it is done indirectly via our choice of null hypothesis (effectively prejudicing the hypothesis of choice).

Hartley thoroughly discusses four overarching paradigms for interpretation of p-values and other items in Chapter 3. These are direct frequentism, indirect frequentism, objective Bayesianism and subjective Bayesianism. Briefly, direct frequentists want to maintain the 'objectivity' of data and the test procedure, and end up alternatively emphasizing a distinction between statistical inference and the formal hypothesis testing procedure, or confusing that distinction. Indirect frequentists explicitly acknowledge that judgment must play a role in interpreting the results of a hypothesis test. The distinction between objective and subjective Bayesians comes down to what types of information/opinion are permitted in forming prior probabilities.

Hartley does a very thorough job of discussing various views and objections expressed in the literature on these topics. I would highly recommend the book based on the critical discussion of these controversies alone. However, this is not the main purpose of the book. Hartley argues that these various views have underlying epistemological foundations. And he claims that these are ultimately religious—in the sense that religious beliefs involve convictions about what is "self-existent" and "self-sufficient."

He also argues that apart from the subjective Bayesian approach, these views presume reductionist "overviews of reality" that create a polarization between two conflicting humanistic ideals, namely the *nature ideal* and the *personality ideal*. Briefly, the nature ideal tends to over-emphasize the certainty of results produced via mathematical procedures, while the personality ideal gives people the freedom to reject or accept ideas as they see fit. These stand in antithesis to one another, but are both usually employed in the practice of statistical inference. But rather than a smooth synthesis, quotes from Ronald Fisher show that a rigidly evidence-based formal process is followed by the freedom on the scientist's part to reinterpret the results as he sees appropriate. Hartley contrasts this with the subjective Bayesian approach that explicitly updates prior knowledge and beliefs using the evidence.

The author concludes that it is best to use the insights of the PLI to ensure that the process of statistical inference makes full use of the advantages of both the frequentist and Bayesian techniques, without succumbing to any of the reductions inherent in taking any one particular paradigm too seriously.