

*Deborah G. Mayo*

## **How Can We Cultivate Senn’s Ability?**

Comment on Stephen Senn, “You May Believe You Are a Bayesian But You’re Probably Wrong”

---

Stephen Senn points out how, in practice, people who claim to have carried out a (subjective) Bayesian analysis have actually done something very different—but that then they heap credit on the Bayesian ideal. “A very standard form of argument I do object to is the one frequently encountered in many applied Bayesian papers where the first paragraphs laud the Bayesian approach on various grounds, in particular its ability to synthesize all sources of information, and in the rest of the paper the authors assume that because they have used the Bayesian machinery of prior distributions and Bayes theorem they have therefore done a good analysis. It is this sort of author who believes that he or she is Bayesian but in practice is wrong.” (58)<sup>1</sup>

Why in practice is this wrong? For starters, Senn points out, the analysis seems to violate such strictures as temporal coherence: “Attempts to explain away the requirement of temporal coherence always seem to require an appeal to a deeper order of things—a level at which inference really takes place that absolves one of the necessity of doing it properly at the level of Bayesian calculation.” (ibid.)

So even if they come out with sensible analyses, Senn is saying, it is despite rather than because they followed strict Bayesian rules and requirements. It is thanks to certain unconscious interventions, never made explicit, and perhaps not even noticed by the Bayesian reasoner. “This is problematic”, Senn thinks, “because it means that the informal has to come to the rescue of the formal”. Not that he thinks there is anything wrong with informality. “Indeed, I think it is inescapable. I am criticizing claims to have found the perfect system of inference as some form of higher logic because the claim looks rather foolish if the only thing that can rescue it from producing silly results is the operation of the subconscious.” (59)

Now, many Bayesians would concede to Senn that in arriving at their outputs they violate strict norms laid down by De Finetti or other subjective Bayesians.

---

<sup>1</sup> Numbers in brackets refer to page numbers of Senn 2011.

But why then do they credit these outputs to some kind of philosophical Bayesianism? The answer, I take Senn to be suggesting, is the fact that they assume that there is but one philosophically righteous position—that of being a Bayesian deep down, where ‘Bayesian deep down’ alludes to a fundamental subjective Bayesian position.

Senn’s idea may be that their belief in Bayesianism deep down is a priori, so it’s little wonder that no empirical facts can shatter their standpoint. (The very definition of an a priori claim is that it’s not open to empirical appraisal.) I think this is generally the case. Many have simply been taught the Bayesian catechism—that subjective Bayesianism is at the foundation of all adequate statistical analyses, and offers the only way to capture uncertainty. Others are true-blue believers (not only in the Bayesian ideal but in the frequentist howlers regularly trotted out). Either way, one can understand why so many Bayesian articles follow the pattern Senn describes: begin by saying grace and end by thanking the Bayesian account for its offer to house all their uncertainties within prior probability distributions, even if in between, the analysis immediately turns to non-Bayesian means that can more ably grapple with both the limits and the goals of the actual inquiry.

Yet Senn, as I understand him, finds this Bayesian routine disingenuous and utterly insufficient as a foundation for statistical research. We ought to be able to look into the black box and recognize that the methods used scarcely toe the (subjective) Bayesian line, or so Senn seems to be saying:

Senn goes on to describe a paper published in *Statistics in Medicine* (Lambert et al. 2005) where the authors consider thirteen different Bayesian approaches to the estimation of the so-called random effects variance in meta-analysis:

“The paper begins with a section in which the authors make various introductory statements about Bayesian inference. For example, ‘in addition to the philosophical advantages of the Bayesian approach, the use of these methods has led to increasingly complex, but realistic, models being fitted’, and ‘an advantage of the Bayesian approach is that the uncertainty in all parameter estimates is taken into account’ (Lambert et al. 2005, 2402), but whereas one can neither deny that more complex models are being fitted than had been the case until fairly recently, nor that the sort of investigations presented in this paper are of interest, these claims are clearly misleading in at least two respects.” (62)

Senn continues:

“First, the ‘philosophical’ advantages to which the authors refer must surely be to the subjective Bayesian approach outlined above, yet what the paper considers is no such thing. None of the thirteen prior distributions considered can possibly reflect what the authors

believe about the random effect variance.<sup>2</sup> [...] Second, the degree of uncertainty must be determined by the degree of certainty and certainty has to be a matter of belief so that it is hard to see how prior distributions that do not incorporate what one believes can be adequate for the purpose of reflecting certainty and uncertainty.” (62f.)

Although Senn at times seems almost to grant that subjective Bayesianism is perfect in theory (or he at least admits to having a love-hate relationship with it), he is clearly “criticising the claim that it is the only system of inference and in particular I am criticising the claim that *because* it is perfect in theory it must be the right thing to use in practice” (59).<sup>3</sup>

Despite these occasional whiffs of being a Bayesian deep down, Senn’s critique would seem to locate him outside the Bayesian (and perhaps any other) formal paradigm. Yet why suppose that this ‘metastatistical standpoint’ admits of no general, non-trivial, empirical standards and principles? It seems to me that one should not suppose this, but instead try and unearth these general arguments, however ‘informal’ or ‘quasi-formal’ they may be. Moreover, I will argue that unless we do so, a Senn-style position here in praise of eclecticism will fail at its intended aim.

Noting that another Bayesian paper a few years later effectively concedes his point, Senn remarks:

“This latter paper by the by is also a fine contribution to practical data-analysis but it is not, despite the claim in the abstract, ‘We conclude that the Bayesian approach has the advantage of naturally allowing for full uncertainty, especially for prediction’, a Bayesian analysis in the De Finetti sense. Consider, for example this statement, ‘An effective number of degrees of freedom for such a t-distribution is difficult to determine, since it depends on the extent of the heterogeneity and the sizes of the within-study standard errors as well as the number of studies in the meta-analysis’. This may or may not be a reasonable practical approach but it is certainly not Bayesian.” (63)

Here, as elsewhere, Senn seems to have no trouble regarding the work as “a fine contribution” to statistical analysis, but one wonders: what criteria is he

<sup>2</sup> He continues: “One problem, which seems to be common to all thirteen prior distributions, is that they are determined independently of belief about the treatment effect. This is unreasonable since large variation in the treatment effect is much more likely if the treatment effect is large (Senn 2007).”

<sup>3</sup> In at least one place Senn slips into the tendency to equate the use of background knowledge to being Bayesian in a subjective sense: Senn declares that a frequentist statistician who chose to set a carry-over effect to zero, in a clinical trial where it fairly obviously warranted being ignored, “would be being more Bayesian in the De Finetti sense than one who used conventional uninformative prior distributions or even Bayes’ factor” (62). (See, in this connection, the discussion in Cox and Mayo 2011 (also *RMM* 2011) on the use of background knowledge.) But there is no evidence that this background knowledge was or needs to be translated into a prior probability distribution.

using to approve it? Is he content to leave those criteria at the unconscious level without making them explicit? If so, isn't he open to the same kinds of subliminal appraisals made by the Bayesians he takes to task? Can we not learn the basis for Senn's sensibility (Senn-sibility)? Does he think that the standards he uses for critically appraising, interpreting, and using statistical methods are ephemeral? Can we say nothing more than that they shouldn't be too terribly awful on any of the four strands of statistical methodology? Senn takes the Bayesian to task for showing us only how to be perfect, but not how to be 'good'. We should move on to this task.

To make this more concrete: How, specifically, would Senn have those authors describe what they actually did, given that it's "certainly not Bayesian"? Now, Senn is not really crediting any overarching or underlying philosophical standpoint for his expertise—but shouldn't he? Is the choice between adopting an a priori standpoint and adopting eclecticism 'all the way down'—even at the level of critically appraising, interpreting, and using statistical methods? If, as Senn himself suggests, most of the Bayesians writing the papers he takes to task are doing what they do more or less unconsciously, then how will he raise their consciousness? Saying it's not really Bayesian doesn't quite tell them what it is.

One might question my presumption that there are some overarching standards, principles, or criteria used in judging work from different schools. But we should at least try to articulate them before assuming it's not possible. And anyway, Senn's remarks suggest he is Senn-sitive to applying a 'second-order' scrutiny.

The account would be far more complex than the neat and tidy accounts often sought: ranging from determining what one wants to learn, breaking it up into piecemeal questions, collecting, modeling, interpreting data and feeding results from one stage into others. Nevertheless, I have suggested there are overarching criteria and patterns of inference based on identifying the error or threat at the particular stage.

To conclude these remarks, then, I want to laud Senn for courageously calling attention to the widespread practice of erroneously describing research as Bayesian, as well as to the tendency of a priori adulation of philosophical Bayesianism Deep Down (BADD). But now that nearly no Bayesians explicitly advocate the one true subjective Bayesian ideal, more is needed. Their position has shifted. While adhering to the BADD ideal, they will still describe their methods as mere approximations of that ideal. After all, they will (and do) say, they can't be perfect, but the Bayesian ideal still lights the way, and therefore discredits all Senn-sible criticism of their claim that all you need is Bayes.

Unless Senn identifies the non-Bayesian work in-between the 'grace and amen' Bayesianism, the worry is that there will be no obligation to amend this practice. Nor is it enough, it seems to me, to merely point out that they are using tools from standard frequentist schools, since these can always be reinterpreted Bayesianly—or so they will say. If it is just a name game, the new-styled Bayesians can say, as some already do about their favorite methods, 'I dub thee

Bayesian’—since ‘Bayesian’ is in the title of my book, or since a conditional probability is used somewhere. That’s the challenge I am posing to those who would advance the current state of statistical foundations.

## References

- Cox, D. and D. M. Mayo (2011), “Statistical Scientist meets a Philosopher of Science”, *RMM* 2, Special Topic: Statistical Science and Philosophy of Science: Where Do (Should) They Meet in 2011 and Beyond? Ed. by D. G. Mayo, A. Spanos and K. W. Staley, 103–114.
- Lambert, P. C., A. J. Sutton, P. R. Burton, K. R. Abrams and D. R. Jones (2005), “How Vague is Vague? A Simulation Study of the Impact of the Use of Vague Prior Distributions in MCMC Using WinBUGS”, *Statistics in Medicine* 24, 2401–2428.
- Senn, S. (2007), “Trying to be precise about Vagueness”, *Statistics in Medicine* 26, 1417–1430.
- Senn, S. (2011), “You May Believe You Are a Bayesian But You Are Probably Wrong”, *RMM* 2, Special Topic: Statistical Science and Philosophy of Science: Where Do (Should) They Meet in 2011 and Beyond? Ed. by D. G. Mayo, A. Spanos and K. W. Staley, 48–66.