Review of: "Straightening the 'Value-Laden Turn': Minimising the Influence of Values in Science"

Michael Evans[1]

1 University of Toronto

Potential competing interests: No potential competing interests to declare.

I'm a statistician and not a philosopher, but I find a lot to agree with in this paper. I was not aware that there was a danger of "values" infiltrating the general scientific enterprise but I think there a few concerns within my own discipline that parallel, and have relevance to, those raised in this paper about the general scientific enterprise.

In my view, and I suppose many others, science is about the search for truth. It is hard, for me at least, to define exactly what is meant by truth but it is intrinsically connected with the concept of objectivity. Objectivity is in itself hard to define precisely although, like truth, it

could be thought of as an ideal that, when we are doing science, we are trying to achieve objectivity as best we can. The concept of objectivity is one that has bedevilled my own discipline where some claim to be objective (frequentists) while claiming that others are subjective (Bayesians) and so unscientific. The Bayesians claim a kind of coherency to their reasoning process that frequentism clearly does not possess so we are left in a bit of

a difficult situation. Actually, neither group is strictly speaking objective, as both approaches involve making choices that are inherently subjective. This is not a problem, however, provided the chosen ingredients can be falsified by checking these against the objective, at least if it is collected correctly, data. So in the end even the Bayesian can claim to be approaching the objective ideal through model checking and checking for prior-data conflict. This issue is important for the paper because it relates to what the author calls "epistemic decisions" as many of the sciences rely on statistical reasoning for this. It is unfortunate, however, that the word "decisions" is used here for a reason I'll now discuss.

Another issue in the field of statistics, and that is relevant to this paper, is that many take the view, whether implicitly or explicitly, that the role of statistics is to make decisions and this involves stating a utility/loss function and then choosing an action (making a decision) based upon finding a statistical procedure that optimizes the expected utility/loss. For example, choosing an estimator of a quantity of interest that minimizes mean-square error or testing a hypothesis by restricting to tests that have an upper bound (the alpha) on the probability of rejecting a hypothesis when it is true and then searching for a test that maximizes the probability of rejecting the hypothesis when it is false, are both applications of decision-theory. There are several problems with this approach not the least of which is that the utility/loss function is a subjective choice (typically not determined by the application) and there is no way to falsify this choice. Another problem with this is that the

evidence in the data may in fact conflict with whatever decision is taken. In other words, the evidence may indicate one answer to the scientific question, e.g., the evidence indicates that a drug alleviates the symptoms of a disease while the decision approach says that the drug is ineffective. In general, this is not a problem as long as we recognize that a proper statistical analysis must always state, as clearly as possible, what the evidence says and, if a decision is taken that contradicts the evidence, then reasons should be provided as to why this is the appropriate decision, e.g., there are too many side effects with this drug or it is too costly to manufacture, etc. I believe this is highly relevant to the distinction the author is recommending in terms of separating values from science, and I fully agree with it. It is to be noted, however, that this distinction is not commonly made in the field of statistics. Unfortunately, the author treats epistemic decisions as accept/reject as this is precisely the decision theory language used in statistics and, as just pointed out, this can conflict with the evidential approach which can present a more appropriate assessment. I think it would be better to discard the language of decision theory and focus on the evidential, sometimes called inferential, approach to statistical reasoning.

So evidence, and in particular statistical evidence, is highly relevant to the discussion in this paper. It is perhaps little understood that the subject concerned with clarifying and making this concept precise, namely, statistics, does not do this in a fully satisfactory way. For example, it is common for applied sciences to use p-values for measuring evidence but it has been known for many years that p-values are not adequate measures of evidence. For example, a p-value cannot indicate evidence in favor of a hypothesis being true, while measuring evidence against a hypothesis, p-values do not measure the strength of evidence and, perhaps most importantly, p-values do not distinguish between scientific and statistical significance. These are just some of the deficiencies that arise from trying to base a reasoning process upon a faulty tool and there are many more. While it is sometimes advocated that the confidence concept can rectify some of the concerns with p-values, a close examination reveals, however, that confidence itself is also rife with difficulties and so is an unsatisfactory solution.

Undoubtedly, these issues play a role in the replication crisis in many fields. After all, if the reasoning process is faulty, how can we expect that there won't be such problems? Perhaps the main problem in statistics as a field, at least in this reviewer's opinion, is the failure of the subject of statistics to precisely define what statistical evidence is, how to measure it and how to use it to determine the subsequent inferences. Since statistical reasoning is a part of much scientific work, from Astronomy to Zoology, it is important that these issues be sorted out. I believe this is highly relevant to the author's thesis as it points to a significant ambiguity in determining "scientific facts".

It this reviewer's opinion that there are approaches to the measurement of statistical evidence that address these issues and that would substantially improve the reasoning processes being used to determine truth in the sciences. For example, this leads to being able to say whether there is evidence against or in favor of a hypothesis and also to a measure of the strength of that evidence. The basic idea is simple and can be referred to
as the \textit{principle of evidence}: starting from prior beliefs (as measured by probabilities) there is evidence in favor if posterior beliefs (obtained by probabilities determined by conditioning on the data) increase belief and there is evidence against if the posterior beliefs are lower than the prior beliefs. This basic idea has interestingly received much more attention in the philosophy of science literature than in statistics which has focused much more on decision theory and optimization. A full description can be found in Evans (2015) which is concerned with the entire enterprise of statistical reasoning for scientific problems. For example, this includes the elicitation of priors, the measurement of the *a priori* bias

in the subjective choices made to effect a statistical analysis and checking all choices made against the objective data to see if the data suggest the choices are not suitable. There is no need for higher order probabilities or for arbitrary scales as mentioned in Section 2.3.2, as measuring the strength of evidence accomplishes the aims of such approaches.

Whether it is this proposal or not, we need an unambiguous and agreed upon statistical reasoning process and I think that fits nicely with the proposals being made in this paper. Values (utilities or losses in statistics) can certainly play a role in what we ultimately do with results of a scientific investigation, but the actual determination of what we believe to be true (always with uncertainties that need to be quantified) has to be as free as we can possibly make it of subjective biases that perhaps reflect what we want to be true. The realization of such a process can only increase real confidence in the conclusions that science draws.

References

Evans, M. (2015) Measuring Statistical Evidence Using Relative Belief. Monographs on Statistics and Applied Probability 144, CRC Press, Taylor & Francis Group.