POST-PRINT
Final version published as:
"Rejoinder"
Franco Taroni, Silvia Bozza, Alex Biedermann, and Colin Aitken
Law, Probability and Risk 2016 15: 31-34
http://lpr.oxfordjournals.org/content/15/1/31.extract?etoc
doi: 10.1093/lpr/mgv007

REJOINDER

F. TARONI, S. BOZZA, A. BIEDERMANN, C. AITKEN

We thank the Editor of *Law, Probability and Risk* for the opportunity to present our paper in the form of a discussion. We also thank the two discussants for their comments and for the time they dedicated to the study of our paper.

We are pleased to see the reactions from such a large group of scientists about a paper that aims to clarify the fundamental term 'uncertainty' and its connection with the use of the Bayes factor as a relevant statistic for evidence evaluation in forensic science. It is encouraging to see that Nordgaard considers the paper "a most important contribution" and that he "agree[s] in principle (...) on the message". However, we will not count on the number of supporters of our view as a measure of the relevance or 'correctness' of the material we have presented. This would amount to *an argumentum ad populum* and which is where Sjerps et al. with their reference to 'other proponents' to the contrary, towards the end of their discussion, seem to lead us.

Let us clarify the issue we have raised: we have asked the question 'what are the components of the likelihood ratio?'. We have seen that these components emphasise the need to consider our probability of the forensic findings, given different propositions. By definition, probability is a number, and this number expresses our uncertainty about a target event, given clearly stated propositions and conditioning information. This personal uncertainty can in principle be elicited, as it exists, by a single number, for you, us and anybody else in their own way. But two things are important. First, we are ***not*** uncertain about this elicited number (only about the event of interest), because it is the number that itself is the expression of uncertainty. Second, there is exactly ***one*** number expressing our state of uncertainty for *two different numbers would express two different states of uncertainty.*

This is a natural understanding of the subjectivist belief-type interpretation of probability, but this choice is not a mere opinion, nor is it adopted by convenience. We advocate the belief-type interpretation of probability and its measurement by reference to a standard because it satisfies an *operational definition* [1], that is a definition that allows us to say precisely what we mean. Its conclusions derive inevitably from the stated assumptions and affect the thinking individual – that is all of us – in a most fundamental way. This may be why people find these conclusions difficult to accept. Yet this view on probability – its measurement by reference to a standard – provides a precise and clear conceptual account of what we mean when we say that we are uncertain about something.

Date: *21 May 2015*.

The upshot of the above is that the denominator of the likelihood ratio is a probability $Pr(\cdot)$, our expression of uncertainty about the findings under stated assumptions, but we are *not* uncertain about the uncertainty expression $Pr(\cdot)$. Hence we are not going to place a probability on a probability − and *hence we have no uncertainty about the likelihood ratio.* This is our main point.

Nordgaard rightly argues, and we fully agree, that there is a difference between the fundamental understanding of probability being given by a single number and our practical proceedings *in the light of* this fundamental understanding: it is what he terms the difficulty of being convinced that one has found one's probability. Thus, the next step should be to find viable practical ways to work with this understanding. This is where the challenge lies, but that challenge is beyond the scope of our paper. However, we are not going to avoid this practical challenge by bending the logic of the theory (of probability) so that it is going to provide a better description of our intuitive perception of our state of uncertainty. We must not confuse the descriptive and the normative perspective, nor should we confuse the concept of uncertainty measurement with the practice of measurement. We entirely share Nordgaard's concerns about the difficulties a forensic scientist might encounter in the assessment of probabilities. Probability assessment in forensic science is a very difficult task for anybody, not only for a scientist whose background is in biology and chemistry and who has not been required to acquire statistical knowledge, but also for a statistician who is unlikely to have the necessary background to understand the principles of forensic investigation. However, as far as the process of evaluation is concerned, we do not believe it should be driven by methodological feasibility, anchored to a forensic scientist's reminiscences from quantitative courses. We agree it is not a major concern of the academic training of forensic scientists to assist statisticians in contrast, it is the statisticians' role to assist the forensic scientist in the process of evaluation [2].

We recognise Nordgaard's observation that in practical proceedings one may end up with several different likelihood ratios, which may be due, for example, to different experts looking at the same findings. Further, we agree with the idea that it is advisable and insightful to consider different values for input parameters and investigate their effect on our likelihood ratio. There is nothing wrong with this technique, known in this context as sensitivity analysis, and a technique which is also supported by the evaluative guideline provided by the European Networks of Forensic Science Institutes (ENFSI). Also, we have never been opposed to the device of framing one's likelihood ratio along expressions such as 'a likelihood ratio in the order of 500 (a million etc.)', acknowledging the above constraints. Indeed, it may be anything from feasible to assign a likelihood ratio as, for example, *exactly* 998. Moreover, such a value is not substantially different from 1000 or 997. All of this is part of established forensic science thinking. None of this, however, involves a procedure to build *a distribution over the likelihood ratio*, an idea which is avoided because it is unsuitable for combination with the belief updating scheme of the recipient of expert information.

On several occasions, Sjerps et al. raise the question of what to report to the court. They emphasise that a population parameter has a probability distribution attached to it.

We could not agree more, as this is the essence of the Bayesian paradigm, and we have never argued the contrary view. We also do not oppose the provision of an opinion on such a distribution *where the object of the reporting is the parameter itself* and a prior can be justified (e.g., in blood alcohol cases, proportion of illicit items in a seizure etc.). This reporting may be a reference to the distribution in its entirety, its credible intervals or convenient summaries. It is reasonable to quantify uncertainty in this context. For example, a legal requirement for alcohol levels can be that the 95% lower bound needs to be above the legal limit for a charge to be brought. We agree with this argument as the concern is with a limit on a measurement. However, the likelihood ratio is concerned with the best measure of the value of the evidence. This is a different concept from a lower limit on a measurement. The determination of a distribution for a parameter is not the object of our discussion. The object of our discussion is the best way in which to report a likelihood ratio. To suggest otherwise would mean a return to times when reporting on DNA profiling results was limited to disparate summaries of population data (so-called 'frequencies'), an approach which is fraught with complications that in the likelihood ratio era can be effectively avoided. In other words, summaries of population data are *not* a summary of the value of evidence.

Sjerps et al. argue that reporting a single number for the likelihood ratio deprives the legal justice system of essential information needed to assess the reliability of the evidence. Highlighting parameter uncertainty is a relevant topic for the discussion of distributions of parameters, but it does not address what Sjerps et al. refer to as reliability. One possible definition of reliability is the question of whether a given likelihood ratio supports the correct proposition. This definition can be addressed by simulating forensic findings given conditions that represent either the prosecution's or the defence case, and obtaining a *single* likelihood ratio in each simulation and hence a distribution. There is no distribution for a likelihood ratio when evaluating the findings in an individual case.

Sjerps et al. also give an example with three experts using different datasets and obtaining the same likelihood ratio. They then state that it is misleading not to tell the courts that the results are not equally reliable. We assume, here, in contrast to the earlier suggested definition, that by reliability Sjerps et al. mean precision, where, statistically, precision is the reciprocal of variance. They then comment that knowledge of reliability is essential information without explaining why it is essential. What is it that is 'essential' about the knowledge of reliability? The likelihood ratio provides the best assessment of the value of the evidence. Knowledge of reliability does not alter that.

Another reason given by Sjerps et al. for a statement of the reliability is that different datasets may give different values for the likelihood ratio. If this is the case, it is a matter to be resolved by cross-examination. It is not a matter to be resolved by considering whether associated probability or confidence intervals overlap. If the different likelihood ratios differ by a large amount (in some sense) this is indicative of differing assumptions or, perhaps, different datasets on the part of the experts. Again, this is a matter to be explored by cross-examination, not by a statement of reliability.

In conclusion, we are *not* going to report on the distribution of the likelihood ratio. Instead, we report on a single value for the likelihood ratio. This is informed by *all* our knowledge of the population parameter, as represented by our posterior distribution for it. The natural way to do this, in the full Bayesian paradigm, *is* to take parameter uncertainty into account by the procedure known as 'extension of the conversation' for discrete data or the integration over the parameter of interest for continuous data [3]. We disagree that integration is a loss of information. Rather, integration is an incorporation of the available information into the evaluation of the evidence in order to provide the best assessment of its value. The result is a single probability assignment that is based on all that we know, at a given instance of time and which is the measure of our uncertainty about the evidence. Once this single number has been determined, it is specific to the case in question. It is not then sensible to attribute uncertainty to this number, with an implicit reference to hypothetical repetitions as in a frequentist procedure.

References

[1] F Lad. *Operational Subjective Statistical Methods : a Mathematical, Philosophical, and Historical Introduction.* John Wiley & Sons, New York, 1996.

[2] DV Lindley, Is our View of Bayesian Statistics too Narrow?, in: JM Bernardo, JO Berger, AP Dawid and AFM Smith (Eds), *Bayesian Statistics 4*, Oxford University Press, Oxford, 1992, pp. 1–15.

[3] N Brummer, A Swart (2014) Bayesian calibration for forensic evidence reporting, arXiv:1403.5997v3 [stat.ML] 10 Jun 2014.