Science and Justice xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Science and Justice

journal homepage: www.elsevier.com/locate/scijus

Reply paper

What should a forensic practitioner's likelihood ratio be? II☆

Geoffrey Stewart Morrison [1]

Forensic Speech Science Laboratory, Centre for Forensic Linguistics, Aston University, Birmingham, England, United Kingdom
Department of Linguistics, University of Alberta, Edmonton, Alberta, Canada
Isaac Newton Institute for Mathematical Sciences, Cambridge, England, United Kingdom

ARTICLE INFO

Available online xxxx
Keywords:
Likelihood ratio
Bayes factor
Precision
Reliability
Accuracy
Validity

ABSTRACT

In the debate as to whether forensic practitioners should assess and report the precision of the strength of evidence statements that they report to the courts, I remain unconvinced by proponents of the position that only a subjectivist concept of probability is legitimate. I consider this position counterproductive for the goal of having forensic practitioners implement, and courts not only accept but demand, logically correct and scientifically valid evaluation of forensic evidence. In considering what would be the best approach for evaluating strength of evidence, I suggest that the *desiderata* be (1) to maximise empirically demonstrable performance; (2) to maximise objectivity in the sense of maximising transparency and replicability, and minimising the potential for cognitive bias; and (3) to constrain and make overt the forensic practitioner's subjective-judgement based decisions so that the appropriateness of those decisions can be debated before the judge in an admissibility hearing and/or before the trier of fact at trial. All approaches require the forensic practitioner to use subjective judgement, but constraining subjective judgement to decisions relating to selection of hypotheses, properties to measure, training and test data to use, and statistical modelling procedures to use – decisions which are remote from the output stage of the analysis – will substantially reduce the potential for cognitive bias. Adopting procedures based on relevant data, quantitative measurements, and statistical models, and directly reporting the output of the statistical models will also maximise transparency and replicability. A procedure which calculates a Bayes factor on the basis of relevant sample data and reference priors is no less objective than a frequentist calculation of a likelihood ratio on the same data. In general, a Bayes factor calculated using uninformative or reference priors will be closer to a value of 1 than a frequentist best estimate likelihood ratio. The bound closest to 1 based on a frequentist best estimate likelihood ratio and an assessment of its precision will also, by definition, be closer to a value of 1 than the frequentist best estimate likelihood ratio. From a practical perspective, both procedures shrink the strength of evidence value towards the neutral value of 1. A single-value Bayes factor or likelihood ratio may be easier for the courts to handle than a distribution. I therefore propose as a potential practical solution, the use of procedures which account for imprecision by shrinking the calculated Bayes factor or likelihood ratio towards 1, the choice of the particular procedure being based on empirical demonstration of performance.

1. Discussion

Much resistance to the adoption of the likelihood ratio framework is not to the idea of assessing the relative probabilities (or likelihoods) of the evidence under prosecution and defence hypotheses per se, but to what is perceived as unwarranted subjective assignment of those probabilities [1,2]. Perhaps wider acceptance will be achieved if greater emphasis is placed on calculation of likelihood ratios via statistical models applied to empirical data and on empirical validation of system performance.

Biedermann, Bozza, Taroni, and Aitken[2] have now made four attempts [3–6] to explain their position in the debate as to whether forensic practitioners should assess and report the precision of strength of evidence statements (likelihood ratios or Bayes factors). Personally, I find the arguments of Biedermann et al. unconvincing because those arguments are based on a premise which a priori I believe to be false, and they have presented no evidence which has convinced me otherwise. The premise is that only a subjectivist concept of probability is legitimate. Under this premise, probability is a state of mind, not a state of nature. The Bayes factor reported by a forensic practitioner is

☆ This paper is part of the Virtual Special Issue entitled: Measuring and Reporting the Precision of Forensic Likelihood Ratios, [http://www.sciencedirect.com/science/journal/13550306/vsi], Guest Edited by G. S. Morrison.

E-mail address: geoff-morrison@forensic-evaluation.net.

[1] The author is also the Guest Editor for the special issue. The present paper was written before the Guest Editor handled any of the other reply papers. The Editor in Chief had editorial responsibility for the present paper, and it was subject to external review.

[2] For simplicity, hereinafter Biedermann et al. Note, however, that in two of the relevant publications the authors are listed in the order Taroni, Bozza, Biederann, Aitken.

http://dx.doi.org/10.1016/j.scijus.2017.08.004
1355-0306/

therefore an expression of their personal belief, and not an estimate of something external to the mind of the forensic practitioner. The Bayes factor that the forensic practitioner reports should therefore be a single value which incorporates all sources of uncertainty affecting their belief. Berger & Slooten [7] take broadly the same position as Biedermann et al.

Biedermann et al. state that assessing the precision of likelihood ratios "involve[s] a misconception of principles and abuse of language" [3], and that the use of non-subjectivist concepts of probability have "arisen from the failure of a scientist to take personal responsibility for their probability assertions" [6]. Insisting that forensic practitioners adopt a subjectivist concept of probability, especially using such confrontational language, is not helpful to the goal of having forensic practitioners implement, and courts not only accept but demand, logically correct and scientifically valid evaluation of forensic evidence. Nordgaard [8] accepts a subjectivist concept of probability, but from a practical perspective argues that it would be counterproductive to force this on forensic practitioners. Martire et al. [9] argue against subjective assignment of probabilities by forensic practitioners, and discuss what is expected and required by the courts.

Even if one believes that, normatively, the trier of fact should act in a subjectivist Bayesian manner, what the court requires from a forensic scientist is not, I suggest, the forensic scientist's subjective opinion, but rather an assessment of strength of evidence based on empirical data and empirically validated procedures. For example, US Federal Rule of Evidence (FRE) 702[3] states that (emphasis added):

> A witness who is *qualified as an expert by knowledge, skill, experience, training, or education* may testify in the form of an *opinion* or otherwise if:
>
> (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
> (b) the testimony is *based on sufficient facts or data*;
> (c) the testimony is *the product of reliable principles and methods*; and
> (d) the expert has *reliably applied the principles and methods to the facts of the case*.

Daubert[4] states that "In a case involving scientific evidence, *evidentiary reliability* will be based upon *scientific validity*." (emphasis in original). That "The adjective 'scientific' implies a grounding in the methods and procedures of science. Similarly, the word 'knowledge' connotes *more than subjective belief or unsupported speculation*" (emphasis added). And that "a key question to be answered in determining whether a theory or technique is scientific knowledge that will assist the trier of fact will be whether it can be (and has been) tested ... '[T]he statements constituting a scientific explanation must be capable of empirical test'". Thus, experience and training constitute criteria for qualifying a forensic practitioner as an expert, but subjective judgement based on experience and training is not sufficient justification for admitting their testimony. I would also suggest that the term "opinion" be read in a restricted sense in which it means that an expert witness may testify as to inferences which they have drawn from facts and data that they observed. FRE 701 states that "If a witness is not testifying as an expert, testimony in the form of an opinion is limited to one that is: (a) *rationally* based on the witness's perception; ..." (emphasis added). FRE 703 states that "An expert may base an opinion on facts or data in the case that the expert has been made aware of or personally observed." And FRE 705 states that the expert witness may be required to state the reasons for their opinion and disclose the facts or data on which is it based. Thus, an opinion does not simply mean whatever a witness believes, but what they can rationally infer from what they have observed. Observation in the form of personal perception for non-expert witnesses and observation of facts or data for expert witnesses. Much of FRE 702 and *Daubert* is then concerned with necessary conditions regarding the process by which experts draw inferences and demonstrate scientific validity (see also the 2016 report by President Obama's Council of Advisors on Science and Technology [10] for commentary on what constitutes scientific validity in the context of FRE 702).

In contrast to the position of Biedermann et al., Sjerps et al. [11] and Morrison & Enzinger [12] argue that once the forensic practitioner has made explicit the prosecution and defence hypotheses that they have adopted, including the relevant population specified as part of the defence hypothesis, and they have made explicit what properties they will measure, then there are true but unknown population[5] distributions and the forensic practitioner's task is to estimate likelihoods from those distributions using models trained on relevant sample data. There are subjective decisions to be made, including selecting hypotheses that are expected to address an appropriate question of interest to the trier of fact, and selecting sample data which are sufficiently representative of the known source and relevant population specified in the prosecution and defence hypotheses. These are pre-empirical decisions which require subjective judgements on the part of the forensic practitioner. This should be made absolutely clear in the case report; first, so that the judge at an admissibility hearing and the trier of fact at trial can consider whether the question the forensic practitioner set out to answer is actually an appropriate question, and whether the data and statistical models used by the forensic practitioner are actually answering that question; and, second, so that the trier of fact can understand the meaning of the likelihood ratio value that the forensic practitioner provides in answer to that question – if one does not understand the question, one cannot understand the answer. The appropriateness of the forensic practitioner's subjective judgements in these matters is something which should be debated by the parties before the judge at an admissibility hearing and/or the trier of fact at trial, in the first instance with respect to admissibility and in the second instance with respect to weight.

The forensic practitioner should also empirically test the performance of their system (measurement and statistical modelling procedures) using test data which represent the relevant population and reflect the known-sample and questioned-specimen conditions. Again, the appropriateness of the test data depends on a subjective judgement made by the forensic practitioner, which ultimately needs to be accepted or rejected by the judge at an admissibility hearing or the trier of fact at trial. If the test data were not sufficiently representative of the relevant population and reflective of the case conditions, then the results of the empirical testing would not be informative as to the validity and reliability of the system when applied to the actual known-source sample and questioned-source specimen in the case. If the judge decides that the test data are appropriate, then the judge can consider whether the demonstrated degree of validity and reliability is sufficient to warrant admission of testimony based on the system that was tested.[6]

The ability of the forensic practitioner to make good subjective judgements on the pre-empirical matters discussed in the last two paragraphs will depend on their expertise gained via training and experience, and these subjective judgements must ultimately be accepted or rejected by the judge and/or trier of fact. If, however, the remainder of the process consists of quantitative measurements and statistical models, and the output of the statistical model is directly reported as the strength of evidence statement, such procedures do not involve additional subjective judgement [13]. The latter procedures are

[3] Federal Rule of Evidence 702 as amended Apr. 17, 2000, eff. Dec. 1, 2000; Apr. 26, 2011, eff. Dec. 1, 2011.

[4] *William Daubert* et al. *v Merrell Dow Pharmaceuticals Inc.*, 509 US 579 (1993)

[5] In this instance I use the word "population" as a contrast with "sample", not to contrast "relevant population" with "known source", hence I am referring to both a relevant-population population distribution and a known-source population distribution.

[6] For extended discussion of the topics covered in the last two paragraphs, see [13–15].

Please cite this article as: G.S. Morrison, What should a forensic practitioner's likelihood ratio be? II, Sci. Justice (2017), http://dx.doi.org/10.1016/j.scijus.2017.08.004