Science and Justice 57 (2017) 472–476

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

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

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].

transparent and replicable, and not susceptible to cognitive bias. 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 – substantially reduces the potential for cognitive bias. Lack of transparency and replicability, and susceptibility to cognitive bias are serious problems for approaches in which the strength of evidence statement is directly the result of a forensic practitioner's subjective judgement [16–20]. The goal of the approach outlined here is not complete objectivity – complete objectivity is impossible. Rather, the goal is to maximise objectivity in the sense of maximising transparency and replicability, and minimising the potential for cognitive bias, and 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. Considering the appropriateness of these pre-empirical subjective judgements, then considering the sufficiency of the empirically demonstrated degree of validity and reliability, should be more manageable tasks than deciding on the merit of a strength of evidence statement which is directly a forensic practitioner's subjective judgement.

For a frequentist, the forensic practitioner's likelihood ratio is an estimate of a true but unknown value, and that estimate should be accompanied by an assessment of its imprecision due to sampling uncertainty. For a subjectivist Bayesian, the forensic practitioner's Bayes factor is their state of belief, and the single value they report should already incorporate all sources of uncertainty. In their latest paper in the current debate [6], Biedermann et al. make it clear that the procedures they advocate do not allow forensic practitioners to directly assign likelihood ratios based only on their subjective judgement without consideration of sample data. Instead, they advocate explicit formal calculation of Bayes factors based on a combination of sample data and prior probability distributions, with integration over nuisance parameters. Once one has specified the hypotheses adopted, the properties measured, and the prior probability distributions, training data, and statistical models to be used, then the procedures for calculating a Bayes factor are as transparent, replicable, and resistant to cognitive bias as the procedures for a frequentist calculation of a likelihood ratio. The only additional elements which must be specified in the Bayesian approach are the prior probability distributions. Since the choice of prior distributions is a subjective judgement on the part of the forensic practitioner, this is also something which should be made explicit and its appropriateness debated before the judge in an admissibility hearing and/or before the trier of fact at trial.

Jaynes [21] advises (p. 373, emphasis in original):

> problems of inference are ill-posed until we recognize three essential things.
>
> (A) The prior probabilities represent our prior *information*, and are to be determined, not by introspection, but by *logical analysis* of that information.
> (B) Since the final conclusions depend necessarily on both the prior information and the data, it follows that, in formulating a problem, one must specify the prior information to be used just as fully as one specifies the data.
> (C) Our goal is that inferences are to be completely 'objective' in the sense that two persons with the same prior information must assign the same prior probabilities.
> If one fails to specify the prior information, a problem of inference is just as ill-posed as if one had failed to specify the data.

Biedermann et al. do not discuss forensic practitioners' choice of prior distributions. I suggest that the best choice would not be prior distributions which are solely the result of a practitioner's subjective judgement based on their training and experience. The forensic practitioner will have to attempt to justify their choice and show that they have taken steps to reduce the potential for cognitive bias. I suggest that this will be much easier if the priors are based on empirical data (raw data or published summary statistics) which the practitioner can argue are relevant (e.g., Morrison et al. [22]), or if the priors are uninformative or reference priors (see Curran [23]).

I present an example which assumes unconstrained continuously-valued univariate data. A datum can have any value, x, between minus and plus infinity. In Fig. 1, the dotted curve represents a relatively uninformative prior distribution – it is a wide flat distribution. The solid curves represent Gaussian distributions fitted to sample data – 8 data points were used to estimate the mean for each of the leftmost and the rightmost distributions and all 16 data point were used to estimate a pooled variance.[7] The dashed curves represent Bayesian posterior predictive probability distributions calculated using both the sample statistics and the hyperparameters of the prior distribution. Both the leftmost and rightmost posterior predictive probability distributions were calculated using the same prior distribution (the one shown as the dotted curve). In calculating posterior predictive probability distributions, the relative weight of the sample statistics and the prior-distribution hyperparameters depends on the number of data points in the sample. If only a small amount of sample data is available, as in this example (and as is often the case in forensic applications), the Bayesian posterior predictive probability distributions will be substantially flatter and wider than the sample distributions. If the amount of sample data is large, the Bayesian posterior predictive probability distributions will approximate the sample distributions.

The result of wider flatter posterior predictive distributions in both the numerator and the denominator is that the Bayes factor will be closer to the neutral value of 1 than the frequentist best estimate for the likelihood ratio calculated using only the sample data. Taking the rightmost distribution in Fig. 1 as the distribution for the numerator of a likelihood ratio and the leftmost distribution as the distribution for the denominator, the y axis in Fig. 2 gives the log base 10 likelihood ratios corresponding to the values on the x axis (Figs. 1 and 2 both have the same x axis). The log Bayes factor values (dashed curve in Fig. 2) are always closer to the neutral value of 0 than the frequentist log likelihood ratio values (solid line in Fig. 2), except trivially when they both equal 0. Taking the exponents, the Bayes factor values are always closer to the neutral value of 1 than the frequentist likelihood ratio values, except trivially when they both equal 1. For some examples of application of this approach see Brümmer & Swart [24] and Zhang et al. [25].

The smaller the amount of sample data, the closer to 1 the Bayes factor calculated using uninformative priors will be, and the worst the precision of the frequentist likelihood ratio will be. From a practical perspective a Bayesian procedure using uninformative priors therefore has the same effect as a frequentist procedure which uses a more neutral value than the best estimate, with the degree of shrinkage towards 1 based on the assessed degree of precision. I refer here to the general effect, not the exact numerical values.

The Bayesian procedure with uninformative priors is no less objective than the frequentist procedure. Use of reference priors (e.g., Jeffreys reference priors [26]), as opposed to uninformative priors in general, could even be considered more objective in that one would not have to make an arbitrary decision as to the degree of shrinkage.

There are other (not necessarily Bayesian) procedures which result in shrinkage. One example is regularized logistic regression ([27] §4.4.5, [28]). Logistic regression is linear in the logged odds domain, and would therefore produce a straight line similar to the solid line in Fig. 2. Regularized logistic regression, however, would result in a line with a shallower slope, and hence would produce log likelihood ratios that are closer to 0 (likelihood ratios that are closer to 1). The extent of shrinkage would be controlled by the size of the regularisation

[7] For illustrative purposes the values of these statistics were specified rather than calculated from actual data.

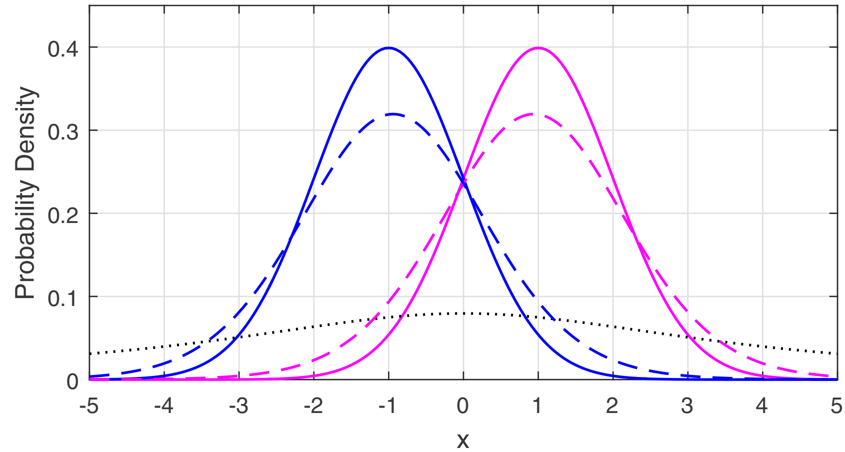

Fig. 1. Dotted curve: relatively uninformative prior distribution. Solid curves: Gaussian distributions fitted to sample data. Dashed curves: Bayesian posterior predictive distributions.

weight, and a disadvantage is that the size of that weight would have to be chosen.

2. Proposal

Curran [29] (p. 380) points out that:

> An astute lawyer would also ask "If I took another sample of size 200, would this figure change?" The single most effective response to this question is "Yes, and my method for assessing this probability has already taken this into account."

And opines that:

> I believe that an expert witness who has used a statistically justifiable method for quantifying and adjusting for sampling uncertainty in his or her evaluation will be well-equipped to respond to the sample size question.

The question is what is the best method for doing this? The impediment to reaching a consensual answer to this question seems to be due to paradigmatic philosophical differences down to the level of the meaning of probability. Is there, however, a practical solution which could be acceptable to all?

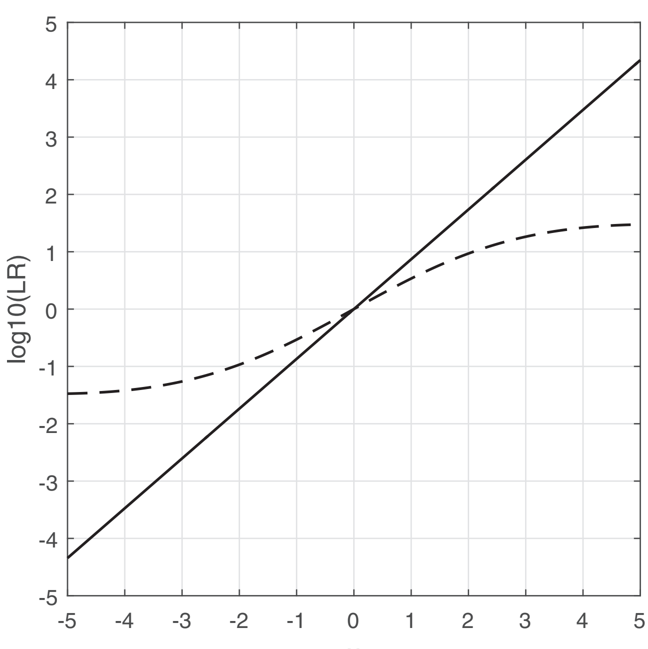

Fig. 2. Solid line: frequentist log likelihood ratio values derived from the Gaussian distributions fitted to the sample data. Dashed curve: log Bayes factor values derived from the Bayesian posterior predictive distributions.

The discussion above leads me to propose a potential practical solution:

Rather than burden the court with the problem of dealing with a posterior likelihood ratio distribution, or a best estimate plus a coverage interval, or a best estimate plus the bound of the coverage interval closest to 1, the forensic practitioner should calculate and present a Bayes factor or likelihood ratio calculated using a procedure which shrinks the resulting value towards 1.

No particular procedure is mandated. Whatever procedure a practitioner chooses, however, it should be transparent and replicable, resistant to cognitive bias, and its use should be justified via empirical testing. The practitioner should choose a procedure which, based on previous empirical testing under relevant conditions, they believe to be sufficiently valid and reliable. The practitioner should make the test results available to the court so that ultimately the court can decide if the analysis system including the chosen procedure is sufficiently valid and reliable.

This proposed solution does not require the forensic practitioner to adopt a subjectivist concept of probability, nor does it preclude the forensic practitioner from having a subjectivist concept of probability.

The proposed solution does not preclude the use of a Bayesian approach with empirically derived informative priors when the forensic practitioner believes that they can justify their choice of data on which they based the prior distributions.

In addition, the precision or sensitivity of the analysis system actually employed should be empirically assessed and reported as part of validation. As per the recommendations in Taylor et al. [30] and Ommen et al. [31], the results of the precision/sensitivity assessment should be an important factor in the debate as to whether the system employed is sufficiently valid and reliable to be admitted, but it should not be suggested that the results of the precision/sensitivity assessment be used to further shrink the reported value of the strength of evidence towards 1.

Acknowledgements

This work was partially funded by a Simons Foundation Visiting Fellowship. The author would like to thank the Isaac Newton Institute for Mathematical Sciences for its hospitality during the program Probability and Statistics in Forensic Science which was supported by EPSRC Grant Number EP/K032208/1. All opinions expressed are those of the author and, unless explicitly stated otherwise, they do not necessarily reflect the opinions or policies of the funding agencies or of any organizations with which he is affiliated.

References

[1] D.M. Risinger, Reservations about likelihood ratios (and some other aspects of forensic 'Bayesianism'), Law Probab. Risk 12 (2013) 63–73, http://dx.doi.org/10.1093/lpr/mgs011.

[2] D. Faigman, A. Jamieson, C. Noziglia, J. Robertson, R. Wheate, Response to Aitken et al on R v T, Sci. Justice 51 (2011) 213–214, http://dx.doi.org/10.1016/j.scijus.2011.09.001.

[3] F. Taroni, S. Bozza, A. Biedermann, C.G.G. Aitken, Dismissal of the illusion of uncertainty in the assessment of a likelihood ratio, Law Probab. Risk 15 (2016) 1–16, http://dx.doi.org/10.1093/lpr/mgv008.

[4] F. Taroni, S. Bozza, A. Biedermann, C.G.G. Aitken, Rejoinder to Sjerps et al and Nordgaard, Law Probab. Risk 15 (2016) 31–34, http://dx.doi.org/10.1093/lpr/mgv007.

[5] A. Biedermann, S. Bozza, F. Taroni, C.G.G. Aitken, Reframing the debate: a question of probability, not of likelihood ratio, Sci. Justice 56 (2016) 392–396, http://dx.doi.org/10.1016/j.scijus.2016.05.008.

[6] A. Biedermann, S. Bozza, F. Taroni, C.G.G. Aitken, The consequences of understanding expert probability reporting as a decision, Sci. Justice 57 (2017) 80–85, http://dx.doi.org/10.1016/j.scijus.2016.10.005.

[7] C.E.H. Berger, K. Slooten, The LR does not exist, Sci. Justice 56 (2016) 388–391, http://dx.doi.org/10.1016/j.scijus.2016.06.005.

[8] A. Nordgaard, Comment on 'dismissal of the illusion of uncertainty on the assessment of a likelihood ratio' by Taroni F., Bozza S., Biederman A. and Aitken C, Law Probab. Risk 15 (2016) 17–22, http://dx.doi.org/10.1093/lpr/mgv006.

[9] K.A. Martire, G. Edmond, D.J. Navarro, B.R. Newell, On the likelihood of "encapsulating all uncertainty", Sci. Justice 57 (2017) 76–79, http://dx.doi.org/10.1016/j.scijus.2016.10.004.

[10] President's Council of Advisors on Science and Technology, Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-comparison Methods, 2016 (https://obamawhitehouse.archives.gov/administration/eop/ostp/pcast/docsreports/).

[11] M.J. Sjerps, I. Alberink, A. Bolck, R.D. Stoel, P. Vergeer, J.H. van Zanten, Uncertainty and LR- to integrate or not to integrate, that's the question, Law Probab. Risk 15 (2016) 23–29, http://dx.doi.org/10.1093/lpr/mgv005.

[12] G.S. Morrison, E. Enzinger, What should a forensic practitioner's likelihood ratio be? Sci. Justice 56 (2016) 374–379, http://dx.doi.org/10.1016/j.scijus.2016.05.007.

[13] G.S. Morrison, R.D. Stoel, Forensic strength of evidence statements should preferably be likelihood ratios calculated using relevant data, quantitative measurements, and statistical models – a response to Lennard (2013) fingerprint identification: how far have we come? Aust. J. Forensic Sci. 46 (2014) 282–292, http://dx.doi.org/10.1080/00450618.2013.833648.

[14] G.S. Morrison, Distinguishing between forensic science and forensic pseudoscience: testing of validity and reliability, and approaches to forensic voice comparison, Sci. Justice 54 (2014) 245–256, http://dx.doi.org/10.1016/j.scijus.2013.07.004.

[15] G.S. Morrison, W.C. Thompson, Assessing the admissibility of a new generation of forensic voice comparison testimony, Columbia Sci. Technol. Law Rev. 18 (2017) 326–434.

[16] D.M. Risinger, M.J. Saks, W.C. Thompson, R. Rosenthal, The Daubert/Kumho implications of observer effects in forensic science: hidden problems of expectation and suggestion, Calif. Law Rev. 90 (1) (2002) 1–56 (http://www.jstor.org/stable/3481305).

[17] M.J. Saks, D.M. Risinger, R. Rosenthal, W.C. Thompson, Context effects in forensic science: a review and application of the science of science to crime laboratory practice in the United States, Sci. Justice 43 (2003) 77–90, http://dx.doi.org/10.1016/S1355-0306(03)71747-X.

[18] B. Found, Deciphering the human condition: the rise of cognitive forensics, Aust. J. Forensic Sci. 47 (2015) 386–401, http://dx.doi.org/10.1080/00450618.2014.965204.

[19] R.D. Stoel, C.E.H. Berger, W. Kerkhoff, E.J.A.T. Mattijssen, I.E. Dror, Minimizing contextual bias in forensic casework, in: K.J. Strom, M.J. Hickman (Eds.),Forensic Science and the Administration of Justice: Critical Issues and Directions 2015, pp. 67–86, http://dx.doi.org/10.4135/9781483368740.n5.

[20] G. Edmond, A. Towler, B. Growns, G. Ribeiro, B. Found, D. White, K. Ballantyne, R.A. Searston, M.B. Thompson, J.M. Tangen, R.I. Kemp, K. Martire, Thinking forensics: Cognitive science for forensic practitioners, Sci. Justice 57 (2017) 144–154, http://dx.doi.org/10.1016/j.scijus.2016.11.005.

[21] E.T. Jaynes, Probability Theory: The Logic of Science, Cambridge University Press, Cambridge UK, 2003.

[22] G.S. Morrison, J. Lindh, J.M. Curran, Likelihood ratio calculation for a disputed-utterance analysis with limited available data, Speech Comm. 58 (2014) 81–90, http://dx.doi.org/10.1016/j.specom.2013.11.004.

[23] J.M. Curran, An introduction to Bayesian credible intervals for sampling error in DNA profiles, Law Probab. Risk 4 (2005) 115–126, http://dx.doi.org/10.1093/lpr/mgi009.

[24] N. Brümmer, A. Swart, Bayesian calibration for forensic evidence reporting, Proceedings of Interspeech 2014, pp. 388–392 (http://www.isca-speech.org/archive/interspeech_2014/i14_0388.html).

[25] C. Zhang, G.S. Morrison, E. Enzinger, Use of relevant data, quantitative measurements, and statistical models to calculate a likelihood ratio for a Chinese forensic voice comparison case involving two sisters, Forensic Sci. Int. 267 (2016) 115–124, http://dx.doi.org/10.1016/j.forsciint.2016.08.017.

[26] H. Jeffreys, Theory of Probability, 3rd ed. Oxford University Press, Oxford UK, 1961.

[27] T. Hastie, R. Tibshirani, J. Freidman, The Elements of Statistical Learning Data Mining, Inference, and Prediction, 2nd ed. Springer, New York, 2009.

[28] G.S. Morrison, A comparison of procedures for the calculation of forensic likelihood ratios from acoustic-phonetic data: multivariate kernel density (MVKD) versus Gaussian mixture model – universal background model (GMM-UBM), Speech Comm. 53 (2011) 242–256, http://dx.doi.org/10.1016/j.specom.2010.09.005.

[29] J.M. Curran, Admitting to uncertainty in the LR, Sci. Justice 56 (2016) 380–382, http://dx.doi.org/10.1016/j.scijus.2016.05.005.

[30] D. Taylor, T. Hicks, C. Champod, Using sensitivity analyses in Bayesian networks to highlight the impact of data paucity and direct future analyses: a contribution to the debate on measuring and reporting the precision of likelihood ratios, Sci. Justice 56 (2016) 402–410, http://dx.doi.org/10.1016/j.scijus.2016.06.010.

[31] D.M. Ommen, C.P. Saunders, C. Neumann, An argument against presenting interval quantifications as a surrogate for the value of evidence, Sci. Justice 56 (2016) 383–387, http://dx.doi.org/10.1016/j.scijus.2016.07.001.