

# **Response paper to**

## **“The likelihood of encapsulating all uncertainty”:**

### **The relevance of additional information for the LR**

Klaas Slooten<sup>a,b,\*</sup>, Charles Berger<sup>a,c</sup>

<sup>a</sup> *Netherlands Forensic Institute, P.O. Box 24044, 2490 AA The Hague, The Netherlands*

<sup>b</sup> *VU University Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands*

<sup>c</sup> *Leiden University, Institute for Criminal Law and Criminology, PO Box 9520, 2300 RA Leiden, The Netherlands*

© <2017>. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/> ↗

\* Corresponding author

#### **Abstract**

In this response paper, part of the Virtual Special Issue on “Measuring and Reporting the Precision of Forensic Likelihood Ratios”, we further develop our position on likelihood ratios which we described previously in Berger et al. (2016) “The LR does not exist” . Our exposition is inspired by an example given in Martire et al. (2016) “On the likelihood of encapsulating all uncertainty”, where the consequences of obtaining additional information on the LR were discussed. In their example, two experts use the same data in a different way, and the LRs of these experts change differently when new data are taken into account. Using this example as a starting point we will demonstrate that the probability distribution for the frequency of the characteristic observed in trace and reference material can be used to predict how much an LR will change when new data become available. This distribution can thus be useful for such a sensitivity analysis, and address the question of whether to obtain additional data or not. But it does not change the answer to the original question of how to update one’s prior odds based on the evidence, and it does not represent an uncertainty on the likelihood ratio based on the current data.

**Keywords:** Likelihood ratio; Subjective probability; Evidence interpretation

## 1. Introduction

In this short paper we will use the opportunity to react to the various response papers that have been published to the papers in the Virtual Special Issue “Measuring and Reporting the Precision of Forensic Likelihood Ratios”. While we largely agree with Dawid [3] and Biedermann et al. [4], we think it is worthwhile to comment on Martire et al. [2].

Martire et al. state that the assignment of probability “is a mental operation subject to all the frailties of human memory, perception and judgment”. While this is undoubtedly true, this is not related to the issue of whether or not to accept probability to be subjective or not. The issue is not about “The presentation of impression, beliefs or ‘guesses’” versus “objectively true results”, or about guessing numbers and claiming they can’t be wrong. It is, in our opinion, about the philosophical interpretation of the relevant probabilities: as subjective probabilities, or as frequencies. Before we proceed, let us emphasize that we agree with the authors on many counts and we elaborate on these topics first.

We completely agree with Martire et al. [2] that an expert assessment of the weight of evidence should be done in a way which is transparent for the recipient. We have argued that conceptually, a likelihood ratio (LR) is a single number [1]. This does not, however, imply that the LR is the only thing which should be communicated to a court by the forensic expert, and we welcome the opportunity to clarify this. For an LR to be meaningful, the hypotheses should be clearly stated, as should be the evidential data that have been taken into account. In order for the results to be open for criticism and challenge, the statistical / probabilistic model that has been used to calculate the LR, the assumptions that have been made, the necessary parameter estimates etc., should also be made available (directly or upon request). The forensic experts should explain their choice of statistical model and the assumptions that they made so that the appropriateness of the model can be judged. Different experts may apply different models, and consequently arrive at different likelihood ratios, even if they considered identical evidence. A likelihood ratio value alone cannot be examined, and therefore the road towards it must also be made explicit.

We also agree with Martire et al. [2] that biases must be avoided in the evaluation process, and that it can certainly be useful to have evidence evaluated by multiple qualified experts. We agree that there are various pitfalls and difficulties and that it is important, as always in science, to be critical and not to forget common sense.

## 2. Subjective probability

An evaluation of a likelihood ratio necessitates the evaluation of two probabilities. In general we think that these probabilities can only be meaningfully interpreted as subjective probabilities. There are several reasons why we take this position, the main one is that we need to be able to deal with probabilities associated with single events that can only with great difficulty be imagined as coming from a repeated experiment. We believe we have no choice but to accept the consequences of the subjectivist interpretation of probability, with its pros and cons. We believe it is the only one that can be used in the general case, which does not mean we think it is a panacea.

An evaluation of evidence almost always involves more than the evidence itself, it requires a model for interpretation and relevant prior probability distributions. When data are scarce, a likelihood ratio will strongly depend on the prior probabilities chosen by the evaluator. We understand that this may be seen as undesirable, but we believe it simply reflects the reality that data interpretation is made within a context, and that the choice of context is not always obvious.

The same argument can be made about the prior odds of the hypotheses with which the LR will be multiplied. It is not always obvious how to produce these odds, but it simply reflects the reality that new evidence is generally added to existing evidence and information.

The position paper describes the idealized situation, arguing that subjective probabilities are the ingredients of an LR, and that LRs involve the integration over all unknown parameters. In order to carry out this integration, one must choose a probability measure describing the prior distribution of the parameters. In our example, this was a Beta distribution. If it is clear how to choose this distribution, or when enough data are available to make the result depend little on the prior distribution (assuming a prior distribution chosen within some class of reasonable prior distributions), then we believe that one should proceed with this integration in order to obtain the LR.

In reality, integration over the parameters may be problematic since it is certainly not always clear how to choose the mathematical integration measure and different choices will then correspond to LRs that can be quite different. However, as more data become available, different experts whose subjective priors are different will arrive at more similar posterior distributions when they incorporate the same data. Take the example where an LR associated with a shared characteristic between an accused person and (the donor of) a trace is to be assigned for the source question. One must then assess the (subjective) probability that the

accused has the characteristic, given that the trace has it, but that the accused is not the donor. In many cases this amounts to giving a subjective probability of observing the characteristic in a person selected at random from the relevant population. As more data become available, this subjective probability will come closer to the frequentist estimate. However, contrary to what is obtained from a frequentist approach, probability statements about the frequency can be made.

Martire et al. see a problem with LR's that "can be radically altered by only a modest change to the data available". They ask whether an LR can encapsulate all uncertainty. Our answer to this is that the LR summarizes our knowledge and is based on our prior information and on the relevant data that we have. If there is more uncertainty about the frequency of a characteristic of interest, then this reduces our knowledge about it and correspondingly weakens the LR. In our previous contribution we have demonstrated how an LR which suffers from the above problem (by using a very small sample from a population) is usually close to one (unless a prior is used that is very far from uniform), and thus of very little value. It thereby encapsulates all the relevant uncertainty referred to above. This is the uncertainty of an expert who operates rationally and it is limited to the question at hand. Whether the expert in a case is indeed operating rationally, making defensible choices and no mistakes, should – as always – be part of the debate. No paradigm can claim to immunize against all human frailties. If the court is uncertain about the competence of the expert, and this uncertainty does not exist in the expert's own mind, then it should be taken into account in the court's own evaluation. We reemphasize that forensic practitioners should therefore clearly communicate what they have done so that the court can examine it, or have a different expert examine it.

### **3. Martire et al. example: different experts with different interpretations**

Martire et al. give an example where two experts have access to the same data, yet use it differently [2]. By construction, they happen to arrive at the same likelihood ratio, but the change in LR due the introduction of new data is different for both experts. The example is meant to illustrate that different experts may obtain a likelihood ratio via different interpretations of the evidence, and that the weight of evidence obtained may be altered differently if new data are disclosed to both experts. While this is certainly an important point, we have identified various properties of this example that in our opinion distract from the argument made by the authors. In our opinion, even this stylized example contains

subtleties that have not been mentioned by the authors. We will discuss this example in detail, because we believe many elements in it deserve further discussion: the concordance between old and new data, the pitfalls associated with making additional assumptions, and the meaning of a change in the LR caused by new data.

Let us first repeat the example. A crime has been committed in Science City (presumably in the US) by an unknown perpetrator (which we shall denote by  $C$ ), who is known to use the spelling ‘colour’. Someone (denoted  $S$ ) is accused of the crime and, while this was not explicitly mentioned, we will suppose  $S$  is of UK origin and uses the UK spelling ‘colour’.

The prosecution and defense hypotheses in this case are  $H_p : S = C$ , and  $H_d : S \neq C$ , respectively, and the available evidence  $E_C$  about the perpetrator  $C$  is that he makes use of the spelling ‘colour’. In addition, we know two properties of the accused: We denote by  $E_1$  the fact that the accused also uses the spelling ‘colour’, and by  $E_2$  the fact that the accused is from the UK. The question is what likelihood ratio to assign for the findings  $E_C, E_1, E_2$  with respect to the hypotheses  $H_p$  and  $H_d$ .

In the example a database is available containing text fragments authored by 9997 individuals. The database contains records of 98 persons of UK origin and 9899 persons of other origin. In a subset of 497 contributors from the database, 3 are known to use the spelling ‘colour’ and 494 to use the spelling ‘color’. Although this was not explicitly mentioned, we assume that the 3 who used the spelling ‘colour’ are of UK origin and the 494 who used the spelling ‘color’ are not of UK origin.

A first expert (expert A) assumes that a person uses ‘colour’ if and only if that person is from the UK such that  $E_1$  and  $E_2$  become logically equivalent, whereas a second expert (expert B) does not make that assumption. Both experts have access to the aforementioned database, but due to the assumption, expert A can make use of much more data than expert B. However, in this example the LR that they obtain in light of their prior probability distribution for the frequency of the characteristic and the data they can process happens to be the same.

In the appendix we show that the LR we need to compute is

$$\frac{P(E_1, E_2 | H_p, E_C)}{P(E_1, E_2 | H_d, E_C)} = \frac{1}{P(E_1 | H_d, E_C)},$$

so all we really need is the probability  $P(E_1 | H_d, E_c) = P(E_1 | H_d)$ , i.e., the probability that the accused uses the ‘colour’ spelling if he is not the perpetrator (in this stylized example we do not distinguish between source and offence level). Expert A assumes logical equivalence between  $E_1$  and  $E_2$  and therefore assumes that this is the same probability as  $P(E_2 | H_d)$ .

First of all, expert A needs to mention and justify the assumption that people use the spelling ‘colour’ if and only if they are from the UK, making  $E_1$  and  $E_2$  logically equivalent. The database described may be supposed not to contradict this claim, but this was not mentioned by Martire et al. [2]. For example, records directly contradicting expert A’s assumption may be sought. But assuming that the availability of a color / colour recording is independent of UK background, it may also be checked whether in a database where 98 out of 9997 persons are from the UK it is likely to obtain 3 ‘colour’ users in a sample size of 497. Let us assume that there is no reason from these considerations to reject the claim of expert A that only people from the UK use the ‘colour’ spelling and that they all do so. Both experts assign  $P(E_1 | H_d)$  as 1/100 based on their identical prior distributions (i.e., a  $\beta(1,1)$  distribution) and their incorporation of the data, using Equation 7 from Ref. [1].

#### 4. Treatment of new data

Now let us consider what could happen when more data are made available. These data may invalidate the equivalence between  $E_1$  and  $E_2$  by containing a record of a person from the UK who uses ‘color’ or a record from someone who uses ‘colour’ but who is not from the UK. In the example, this does not happen, and 50 persons from the UK using ‘colour’ as well as 50 others, not from the UK and using ‘color’ are added. Martire et al. show that in that case, expert A updates his LR to 67 and expert B updates his LR to 11. The authors warn that “in the absence of explicit information from the expert accurately describing the personal statistical model they applied, the fact finder cannot anticipate how easily or how much an expert might be inclined to change their testimony in light of additional evidence [...] these factors are relevant to a court even if it is not captured by the LR”. We are in partial agreement with this statement. The expert’s LR addresses the question of how to update one’s belief in the hypotheses based on prior knowledge and the current evidence. The impact of possible future evidence is a relevant but separate question. It does not have an impact on the LR that we have based on our current data, but – as we will explain below – it may be useful to decide whether or not to gather additional data, based on our expectation of how much this will alter the LR. We will argue below that experts can quantify how much they

expect their opinion to change if new data would become available, assuming that these data conform to the expert's current expectation based on the present knowledge. Additionally, we believe that a large change of LR for the same evidence due to new data is not as likely to happen as the authors argue.

In this example, it should have been noted first and foremost that these new data are extremely unexpected by both experts. For both experts the rate went up from less than 1% in the previous data to 50% in the new data. It is all but impossible to obtain such new data, if we assume that the original and the new sample are taken from the same population. Either starting from a previous sample containing 98 individuals with a UK origin and 9899 without, or starting from a previous sample with 3 'colour' and 494 'color' users, both experts should remark that the original and new data are so unlikely to have come from the same population obtained via the same sampling strategy, that this alone is reason not to carry out any analysis before that issue is clarified.

Based on the original data we need only really consider the possibilities that the 100 new records contain up to 4 UK residents, and to avoid a contradiction with the assumption of expert A, we will assume that these are exactly the persons using the 'colour' spelling. According to whether there are 0, 1, 2, 3 or 4 UK residents in the new sample (who are the same as the 'colour' users) the likelihood ratios offered by expert A will become 101, 100, 99, 98, 97 respectively, whereas the LRs of expert B will be 120, 100, 86, 75, 67 respectively. We thus see that the LR from expert A is less sensitive to the arrival of new data, which is logical since more data have been used. However, the changes are not dramatic for either expert.

In our opinion, the example mainly illustrates the – difficult – question as to which assumptions are justified. It may be beneficial to make additional assumptions such as expert A had done. Making additional assumptions may allow more data to become relevant, and therefore lead to a LR that is less dependent upon prior (pre-data) distributions. However, any such assumption should be made explicit and justified clearly so that it may be challenged. The fact that opinions may change in the light of new data is a basic requirement of rational interpretation, and not a drawback.

## **5. Interpretation of the frequency distribution of the shared characteristic**

In the example in [1] where an LR is associated with a shared characteristic, we worked out the situation where one starts with a (purely subjective)  $\beta(1, 1)$  prior. Obtaining  $m$  samples

with, and  $n$  without the characteristic updates our distribution to a  $\beta(m + 1, n + 1)$  distribution. The expectation and variance of this distribution determine our LR, which is equal to  $(m + n + 3)/(m + 2)$ .

One may wonder what would be the meaning for the forensic scientist – and possibly a court – of the distribution  $\beta(m + 1, n + 1)$ . In the subjective belief context, this distribution contains the forensic scientist's degree of belief in the event that the next observed person will have the characteristic, but moreover it contains the degree of belief in any possible set of future outcomes, such as, for example, the subjective probability that in the next set of 10 individuals, the characteristic will be observed in three of them. As another example, suppose we have a  $\beta(x, y)$ -distribution, then among the next two individuals there may be 2, 1 or 0 with the characteristic. We know which probability each of these events has in our model and also what LR we would assign for the shared characteristic had we made these observations. These will be either  $(x + y + 3)/(x + 3)$ , or  $(x + y + 3)/(x + 2)$ , or  $(x + y + 3)/(x + 1)$ . One can show that in expectation, the updated LR will be larger than the original one based on  $\beta(x, y)$ . A sketch of the proof can be found in the Appendix.

## 6. Conclusion

We can see several things from this example. The first one is that in expectation it pays to obtain more information in the sense that we expect an LR further away from one. This is because, if we have more data, we expect the frequency of the characteristic to remain the same but it will be based on more data and hence the probability distribution modelling the frequency will be more concentrated around its mean and have a smaller variance. Of course the LR only increases in expectation, we cannot be certain that this will happen: that would be in contradiction with the fact that the LR summarizes our knowledge based on our current data.

Second, we see that we can predict, based on our current belief, how much we expect our belief to change if new data are made available. Therefore, we can also use the distribution  $\beta(x, y)$  to make predictions on the population frequency of the characteristic that we expect to obtain from more data. In other words, the distribution can be used to quantify the additional value we expect to obtain from acquiring more data. We think that the difference between this statement, and uncertainty on the LR might seem subtle but is important. The LR summarizes the evidential weight based on our current data (and prior subjective belief about parameters involved) whereas the distribution of the frequency of the



characteristic quantifies how much we expect the LR to change in the light of new data. Of course, it still remains the case that our expectation of what new data will look like, is determined from a posterior distribution that is itself based on a subjective prior and the currently available data.

## Appendix A

To compute the posterior odds

$$\frac{P(H_p | E_C, E_1, E_2)}{P(H_d | E_C, E_1, E_2)}, \quad (1)$$

we may first write them (by using Bayes' theorem) as

$$\frac{P(E_2 | H_p, E_C, E_1)}{P(E_2 | H_d, E_C, E_1)} \cdot \frac{P(H_p | E_C, E_1)}{P(H_d | E_C, E_1)}. \quad (2)$$

Note that the first quotient is equal to one for both experts: the probability that the accused is from the UK, given that he uses the 'colour' spelling, does not depend on him being the perpetrator. Expert A will set both probabilities in the quotient to one, but expert B, while not doing this, would also obtain a quotient of one because both probabilities are equal. Therefore the required posterior odds can be written as

$$\frac{P(E_1 | H_p, E_C)}{P(E_1 | H_d, E_C)} \cdot \frac{P(H_p | E_C)}{P(H_d | E_C)}. \quad (3)$$

In this product, since we do not condition on knowledge about the accused, the last term is equal to the prior odds  $P(H_p) / P(H_d)$ . Therefore, the relevant likelihood ratio to compute is

$$\frac{P(E_1 | H_p, E_C)}{P(E_1 | H_d, E_C)} = \frac{1}{P(E_1 | H_d, E_C)}. \quad (4)$$

To show that, in expectation, the LR increases when we gather more data, consider that we have a  $\beta(x, y)$ -distribution for the frequency of a shared characteristic. We may have  $x = m + a$ ,  $y = n + b$ , based on a prior  $\beta(a, b)$  distribution and a sample in which we have  $m$  items with the characteristic and  $n$  without. If a trace and a suspect both have the

characteristic, our LR in favor of the suspect having left the trace versus a unknown, random member of the population having left the trace is equal to  $\frac{1+x+y}{1+x}$ , as we have derived in [1].

Now we will compute the LR we would expect if we would add one more item to the database. Based on our current distribution, our probability for the next item having the characteristic is  $x/(x+y)$ , and of course this means our probability that it does not have the characteristic is  $y/(x+y)$ . If we find that the next item does have the characteristic, we update our distribution for the frequency of the characteristic to  $\beta(x+1, y)$ , and otherwise to  $\beta(x, y+1)$ . The respective LRs we compute if, after having inspected the additional item, we observe this characteristic in both the trace and suspect, are

$$\frac{1+x+1+y}{1+x+1} \quad \text{and} \quad \frac{1+x+y+1}{1+x}. \quad (5)$$

In expectation, the LR we will get is therefore

$$\frac{x}{x+y} \cdot \frac{2+x+y}{2+x} + \frac{y}{x+y} \cdot \frac{2+x+y}{1+x}. \quad (6)$$

If we subtract  $\frac{1+x+y}{1+x}$ , the LR based on  $\beta(x, y)$  from this expression, we find, after some

algebra, that this difference equals  $\frac{2y}{(1+x)(2+x)(x+y)}$ , which is a positive number. This

means that the LR we expect if we enlarge the database will always be strictly more than the LR based on the current data. However, we reiterate that this is only in expectation. When we do obtain more items it is also possible that the LR will decrease compared to its value without the additional items.

Intuitively, it is easy to understand that the LR behaves in this way. Indeed, the LR summarizes our knowledge, and so it makes sense that adding data conform our current knowledge and expectation leads to stronger evidence on average (an LR further away from 1). This is because with more knowledge one expects to be able to better distinguish the hypotheses from each other. On the other hand, if we were certain that the evidence would become stronger when based on more data, then we would not have adequately summarized our current knowledge. So it must always be possible that new and less expected data will make the evidence less strong.

## References

- [1] C.E.H. Berger, K. Slooten, The LR does not exist, *Science and Justice* 56 (2016) 388–391.
- [2] K.A. Martire, G. Edmond, D.J. Navarro, B.R. Newell, On the likelihood of “encapsulating all uncertainty”, *Science and Justice* 57 (2017) 76–79.
- [3] A.P. Dawid, Forensic likelihood ratio: Statistical problems and pitfalls, *Science and Justice* 57 (2017) 73–75.
- [4] A. Biedermann, S. Bozza, F. Taroni, C.G.G. Aitken, The consequences of understanding expert probability reporting as a decision, *Science and Justice* 57 (2017) 80–85.