# On Reporting Likelihood Ratio's of Exhaustive and Non-Exhaustive Hypotheses about Rare Events in Criminal Cases

Anne Ruth Mackor[1,*,†], Henry Prakken[2,†]

[1]*Faculty of Law, University of Groningen, Groningen, The Netherlands*
[2]*Department of Information and Computing Sciences, Faculty of Science, Utrecht University, Utrecht, The Netherlands*

### Abstract
This paper discusses how a Bayesian analysis in case of evidence of a rare event (so all its causes are also rare) can be presented in court in more or less misleading ways. Two case studies are used to hypothesise that a method that puts the rarity of both considered causes in the prior odds is less prone to committing probabilistic reasoning fallacies than a method that expresses one of these rarities in a likelihood ratio. Moreover, the second case study is used to argue that when experts report on two hypotheses that are not logically exhaustive, they should either explain why they can still be treated as exhaustive or explicitly warn that no posterior probabilities of the hypotheses can be derived.

## 1. Introduction

In this paper we discuss how probabilistic evidence can be presented in court by experts in clear or misleading ways. We will in particular focus on situations in which the evidence is a rare event so all its possible causes will also be rare. In such cases the 'guilt' hypothesis is often compared to an 'innocence' hypothesis that essentially amounts to coincidence. A fallacy that people then sometimes commit is 'the probability to find this evidence in case of coincidence is so small, this cannot be coincidence, so there must be some other cause of the evidence' (which usually is equated with the guilt hypothesis). Our aim is to provide recommendations on presenting probabilistic evidence in such a way that the probability that fact finders commit this fallacy is reduced. We will approach this issue by way of a discussion of two cases: the Sally Clark case in the UK [1] and a recent Dutch case involving a large number of car collisions on car parks all involving the same drivers[1].

After a brief introduction of the basic concepts of Bayesian probability theory we first summarise two ways in which Dawid [1] analyses the Sally Clark case. We then present our own analysis of how expert witnesses reported likelihood ratio's in the serial car collision case and discuss relevant similarities between this case and the Sally Clark case. In particular, in both cases the evidence is a rare event so all its possible causes will also be rare. We will hypothesise that a representation method in which the rarity of both considered causes is presented in the prior odds is potentially less misleading than a method in which the rarity of one of these events is instead represented in the likelihood ratio (as was done in the serial car collision case).

We then discuss a second problem with the way the probabilistic evidence was presented in the serial car collision case, namely, the fact that the hypotheses considered by the forensic expert are, unlike in the Sally Clark case, cannot be reasonably regarded as exhaustive. We argue that a forensic expert cannot report a likelihood ratio about such hypotheses without making it very clear that when the hypotheses are not exhaustive, the court cannot conclude anything about their posterior probability.

---

[1]ECLI:NL:RBOBR:2024:3277 and ECLI:NL:GHSHE:2025:2019.

## 2. Basics of Bayesian probability theory

In this section we review the basics of Bayesian probability theory as far as necessary for our purposes. As for notation, $Pr(A)$ stands for the unconditional probability of $A$ while $Pr(A \mid B)$ stands for the conditional probability of $A$ given $B$. In criminal cases we are interested in the conditional probability $Pr(H \mid E)$ of a hypothesis $H$ of interest (for instance, that the suspect is guilty of the charge) given evidence $E$ (where $E$ may be a conjunction of individual pieces of evidence). For any statement $A$, the probabilities of $A$ and $\neg A$ add up to 1, as do the conditional probabilities $Pr(A \mid B)$ and $Pr(\neg A \mid B)$ for any $B$. The same holds for hypotheses $H_1$ and $H_2$ that do not logically negate each other but that still exclude each other (they cannot both be true) and are exhaustive (no other hypothesis can be true) on other grounds. Consider, for instance, two hypotheses $H_1$ that person A is the (sole) perpetrator and $H_2$ that another person B is the perpetrator. These hypotheses are mutually exclusive but they do not logically negate each other. Nevertheless, if there is evidence that only $A$ or $B$ can be the perpetrator, then $H_1$ and $H_2$ can still be reasonably assumed to be exhaustive.

We consider the often occurring situation that a forensic expert reports on the relation between two hypotheses $H_1$ and $H_2$ and a single piece of evidence $E$. Bayes' theorem then becomes (in odds form)

$$\frac{Pr(H_1 \mid E)}{Pr(H_2 \mid E)} = \frac{Pr(E \mid H_1)}{Pr(E \mid H_2)} \times \frac{Pr(H_1)}{Pr(H_2)}$$

In words, the posterior odds of $H_1$ and $H_2$ equals their likelihood ratio times their prior odds. Here $H_1$ and $H_2$ are mutually exclusive but not necessarily exhaustive. When more evidence is available, its likelihood ratio can (under the appropriate assumptions of statistical independence) be used in a new application of Bayes' theorem in which the posterior odds given the initial evidence $E$ is used as the prior odds.

## 3. Dawid on the Sally Clark case

The Sally Clark case is a tragic case that happened in England. In December 1996 Sally Clark's first son suddenly died, 2,5 months old, while he was alone at home with his mother. In January 1998 Sally's second son died, 2 months old, also while being at home alone with his mother. Sally was accused of having killed her sons but Sally claimed they had died of natural causes (such as Sudden Infant Death Syndrome, also known as cot death). A paediatrician estimated the probability that one child dies from unexplained natural causes in a family such as the Clarks is 1 in 8500. He then multiplied this probability with itself to conclude that the probability that two children die from unexplained natural causes in a family such as the Clarks is 1 in 73 million. Many may be tempted to infer from this that Sally almost certainly killed her two sons and indeed the jury found Sally Clark guilty and her first appeal was dismissed. However, this inference is based on at least two probabilistic reasoning errors.

Dawid, in an expert report for the appeal case and in [1] convincingly showed that the expert should not have multiplied the 1 in 8500 probability with itself since it cannot be assumed that two deaths from unexplained natural causes in the same family are statistically independent of each other. We will therefore assume an arguably better founded probability of of 1 in 850,000 that two children die from unexplained natural causes in a family such as the Clarks[2].

Regarding the second reasoning error, Dawid discusses two ways to model the case with the odds form of Bayes' theorem. In one of them he considers the hypotheses $H$ that Sally Clark killed her two babies and $\neg H$ that she did not kill her two babies. Note that $\neg H$ leaves open the possibility that the babies did not die at all. Bayes' theorem is thus instantiated as follows:

$$\frac{Pr(H \mid E)}{Pr(\neg H \mid E)} = \frac{Pr(E \mid H)}{Pr(E \mid \neg H)} \times \frac{Pr(H)}{Pr(\neg H)}$$

which with the 1 in 850,000 probability becomes

---

[2] Even higher probabilities have been estimated. See e.g. [2] or https://plus.maths.org/content/beyond-reasonable-doubt.

$$\frac{Pr(H \mid E)}{Pr(\neg H \mid E)} = 850,000 \times \frac{Pr(H)}{Pr(\neg H)}$$

The reason that the likelihood ratio equals 850,000 is that if Sally Clark killed her two babies ($H$), they surely died ($E$). So the death of Sally Clark's two sons is strongly incriminating evidence. However, the prior odds counters this strength, since the probability that Sally Clark killed her two babies is also very low: there are not many mothers who kill their children. Let us assume for ease of calculation that the prior probability that Sally Clark killed her babies is 1 in 1.7 million. This yields a prior odds of almost 1 in 1.7 million, so the posterior odds is $\frac{1}{2}$, which yields a posterior probability that Sally Clark killed her two sons of just 33.3%.[3]

A crucial observation here is that the probability of two rare events must be compared: not only unexplained death of two babies by natural causes is rare but also a mother killing her two baby sons is rare. In the above Bayesian modelling the rarity of the first event is accounted for in the likelihood ratio, which is high, while the rarity of double murder is expressed in the prior odds, which is low.

In legal practice a problem may arise if an expert only reports the likelihood ratio of 850,000, which implies that the probability of the two deaths given unexplained natural causes is very low. A person not trained in probability theory could easily infer from this that the death of the two children cannot be due to natural causes, so Sally Clark must have killed her babies. However, this is the well-known fallacy of transposing a conditional probability [3, 4].

The risk of committing this fallacy does not occur in Dawid's alternative Bayesian modelling of the case, in which he considers the hypotheses $H_1$ that the two babies died since Sally Clark killed them and $H_2$ that the two babies died from unexplained natural causes. Note that these two hypotheses, although mutually exclusive, are not exhaustive. For instance, one of the babies could have been killed by Sally Clark while the other died from natural causes, or someone else than Sally Clark could have killed the two babies. Yet Dawid treats the hypotheses as exhaustive. In our opinion, this can in general be justified, albeit defeasibly, on the basis of the available evidence in the case (if it does not point to any other possible hypothesis) and/or on the basis of background knowledge. (See also [5].) For example, there could be evidence that only Sally Clark could have killed the babies, or that the babies had no known diseases. That such an assumption of exhaustiveness is defeasible means that it can be invalidated by further evidence. We leave it to the reader to assess whether Dawid's assumption of $H_1$ and $H_2$ as exhaustive can be defeasibly justified.

We now again assume that the prior probability of $H_1$ equals 1 in 1.7 million. The assumption of exhaustiveness of $H_1$ and $H_2$ must be expressed by letting their prior probabilities add up to 1. This yields a prior odds of 0.5 since a frequency of 1 in 850,000 is twice the frequency of 1 in 1.7 million. Moreover, the likelihood ratio of the evidence (that the two babies died) relative to $H_1$ and $H_2$ is now 1 since both hypotheses now imply the evidence. Dawid's alternative modelling then becomes:

$$\frac{Pr(H_1 \mid E)}{Pr(H_2 \mid E)} = 1 \times \frac{1}{2} = \frac{1}{2}.$$

So under the assumption of exhaustiveness of $H_1$ and $H_2$ this again yields a posterior probability that Sally Clark killed her two sons of just 33.3%.

Although both methods thus lead to the same results, they can still make a difference in practice. As we noted above, the first representation method can cause fact finders to fallaciously transpose the conditional if they are not warned against this fallacy. By contrast, in the second method the fact finders are actively stimulated to consider the relative rarity of the two hypotheses. Accordingly, we hypothesise that the second method is less misleading and should therefore be preferred over the first method.

In the next section we will discuss an example in which at least one court may have been been misled by a report using the first method.

---

[3] We do not claim that this is a well-founded posterior probability. We have chosen the various numbers since they seem not unreasonable given the literature on the case (e.g. [2] and since they can be used to illustrate the fallacy of the transposed conditional.

## 4. The serial car collision case

We next discuss a recent Dutch case of a series of car collisions. A couple was during a period of 74 months involved in 56 car collisions that happened at various car parks in the Netherlands (below this is evidence $E$). They were prosecuted for insurance fraud by deliberately having caused these collisions. Their defence was that the collisions were unintentionally caused since they were poor drivers. An expert of the Dutch forensic institute (NFI) compared the following two hypotheses, in which a 'systematic cause' is in the expert report essentially defined as the logical negation of 'accidental'.

- $H_1$: the suspects have a normal risk of accidental collisions, and a large number of the collisions involving the suspects have a systematic cause.
- $H_2$: the suspects have an extremely high risk of coincidental collisions, and all collisions involving the suspects are coincidental.

The expert then estimated the likelihood ratio $\frac{Pr(E|H_1)}{Pr(E|H_2)}$ as higher than 1 million. This number was computed as follows. First $Pr(E \mid H_2)$ was on the basis of data from insurance companies about claims concerning car collisions determined as less than 1 in a million. This number seems reasonable. Then $Pr(E \mid H_1)$ was computed as follows. First, on the basis of the data the probability of 0 coincidental collisions given $H_1$ was determined as 0,91. This also seems reasonable. The expert then concluded from this that *therefore all collisions were accidental, therefore $Pr(E \mid H_1) = 0,91$.* At first sight, this would seem to be a reasoning fallacy, but it turns out that the expert assumed in $H_1$ that 56 *intentional* collisions had happened. And then the probability that exactly 56 collisions happened given $H_1$ indeed equals the probability of 0 coincidental collisions given $H_1$.

Although thus the expert's analysis is mathematically correct, we believe that the presentation is highly misleading, since the report gives no indication whatsoever that the expert has in $H_1$ assumed that 56 intentional collisions had happened. Yet it is crucial for judges to know this, since under this assumption the prior probability of $H_1$ greatly reduces, just as in Dawid's first method of representing the Clark case (see Section 3 above). After all, very few normal drivers will intentionally cause 56 collisions in 74 months, even if a large number of the collisions they do cause is intentional. Although the report contains a warning that the likelihood ratio must be combined with the prior odds, it says nothing about the above mentioned assumptions and the resulting specific compensating effect on the prior odds.

The suspects were convicted, after which they appealed and were convicted again. They appealed with the Dutch Supreme Court, which case has not yet been decided. In the initial trial[4] the court justified its conviction in part by referring both to the NFI report and a study (from another case) of the Dutch Association of Insurers, which considered 15 car accident during a period of 9 years and concluded:

> The probability that a specific driver has by coincidence become the victim of 15 car collisions during a period of 9 years is negligibly small. Moreover, the probability that 15 car collisions during a period of 9 years are due to coincidence is many time smaller. The accidents must therefore be due to other causes than coincidence.

This clearly is an instance of the fallacy of transposing the conditional and the fact that the court refers to it as support for its conviction indicates that the court has also become victim of this fallacy. Does this mean that this case is a miscarriage of justice? Not necessarily, since there was also other evidence in the case, notably witness statements that the *modus operandi* was in all collisions the same and indicated the intention to cause the collisions. It may well be that this evidence, when combined with the evidence as considered above, leads to a high posterior odds of hypothesis $H_1$ versus $H_2$ given all considered evidence. Incidentally, the text of the appeal case[5] gives no indications whether the court

---

[4] ECLI:NL:RBOBR:2024:3277.
[5] ECLI:NL:GHSHE:2025:2019.

of appeal committed a similar fallacy. In any case, it does not refer to the study of the Dutch Association of Insurers.

We believe there is a second problem with the way the NFI expert reported the likelihood ratio. In Section 3 we argued that when two considered hypotheses are not each other's logical negation, they can possibly still reasonably (albeit defeasibly) be assumed to be exhaustive on the basis of the available evidence and/or background knowledge. However, in the car collision case such an assumption does not seem warranted. The NFI report does not formulate the 'innocence' hypothesis as the logical negation of the 'guilt' hypothesis (unlike in Dawid's first method). If, by contrast, this is done, then the prior odds may well greatly decrease. For instance, the negation of $H_1$ includes the possibility that the suspects are normal drivers whose collisions are all coincidental. It is true that $Pr(E|\neg H_1)$ will be higher than $Pr(E|H_2)$, but it is not obvious that this would fully compensate the decrease of the prior odds. Hence without further justification there seems no reasonable assumption under which the exhaustiveness of the two hypotheses can be defeasibly justified. Accordingly, their prior probabilities do not even defeasibly add up to 1. But this implies that the posterior probability of the guilt hypothesis $H_1$ cannot be determined: the posterior odds of $H_1$ and $H_2$ cannot be translated in the posterior probability of guilt or innocence since another hypothesis than $H_1$ or $H_2$ may be true.

In our opinion forensic experts have to report their probabilistic evidence in a logically correct and complete way, which means that they should at least inform the fact finder that if the considered hypotheses cannot be reasonably regarded as exhaustive, the fact finder cannot draw any conclusion about their posterior probability. Moreover, we believe that forensic experts should preferably not report about such hypotheses at all, in order to minimise the risk of incorrect interpretations and probabilistic reasoning fallacies.

The NFI forensic expert did not report on the non-exhaustiveness of the considered hypotheses and the implications thereof. Moreover, there is no indication in the ruling of the court of appeal that it fully grasped the relevance and the implications of the non-exhaustiveness. Finally, we note that the general tutorial text of the NFI on probability theory to which the expert refers in her report[6] says nothing about the implications of reporting likelihood ratio's of non-exhaustive hypotheses. We therefore conclude that it cannot be excluded that the NFI report had a misleading effect on the courts in the initial and/or appeal trial.

## 5. Recommendations

In this section we make some recommendations on the basis of our discussions. Note first that, as shown by Dawid in [1, 6] the choice between the two methods to report likelihood ratio's considered in Section 3 is mathematically speaking arbitrary. This in turn implies the same for the often-applied policy of forensic experts to only report likelihood ratio's. This policy is motivated by the argument that determining the prior would be outside the expert's expertise and therefore the task of the fact-finder (see e.g. [5, p. 2] or p. 5 of the NFI's general text mentioned above). However, the mathematical equivalence of both representation methods implies that if the expert can say something about the likelihood ratio in the first method (with hypotheses that logically negate each other) then the expert is also be able to say something about the prior of coincidence in the second method (with the evidence incorporated in one or both hypotheses). Moreover, we conjecture that in cases with rare events, experts will on the basis of their expertise sometimes also be able to say something about the prior of the other hypothesis, especially if both of the prior probabilities can be determined on the basis of experiments or statistical data analysis.

Accordingly, we make the following recommendations.

1. In cases with rare evidence all its potential causes will also be rare. Therefore, likelihood ratio's should in such cases preferably be reported in the second of Dawid's methods, in which the rarity of both considered causes is represented in the prior odds. In the serial car collision case

---

[6]https://www.forensischinstituut.nl/publicaties/publicaties/2017/10/18/vakbijlage-waarschijnlijkheidstermen (in Dutch)

this would amount to letting both hypotheses imply that the suspects experienced 56 accidents. Moreover, experts should in such cases, if possible, also say something about the prior odds.

2. If instead a guilt hypothesis is chosen which makes assumptions resulting in a high likelihood ratio but in a low prior odds, then this should be very explicitly reported to the fact finder.

3. In general, experts should not report on hypothesis pairs that cannot reasonably be assumed to be exhaustive. Moreover, in cases in which hypothesis pairs that can be reasonably but only defeasibly be assumed to be exhaustive, which means that further evidence may invalidate the assumption, the consequences of this should be clearly explained to the fact finders.

As regards our first recommendation, sometimes another reason is given why the first method, with logically negating hypotheses, is suboptimal (see e.g. [5, 7], namely, that it is often hard to determine the probability of the negation of the guilt hypothesis since it does not correspond to a well-defined specific event. Granted that this is true, it still does not imply that hypotheses should at least on reasonable (though defeasible) assumptions be exhaustive, since otherwise it still holds that their posterior probability cannot be reasonably determined.

## 6. Conclusion

In this paper we analysed how in two cases forensic experts have reported likelihood ratio's about rare evidence. We observed that in such cases all possible causes of the evidence will also be rare, which is sometimes not understood by fact finders. As a consequence, fact finders are prone to committing the fallacy of the transposed conditional if the meaning and implications of the probabilistic evidence are not very clearly explained to them. We hypothesised that the best representation is to use Dawid's second method by choosing hypotheses that both imply the rare evidence, so that the relative rarity of the considered hypotheses is explicit in the prior odds.

We then noted that in the serial car collision case the two hypotheses considered by the forensic experts could not be reasonably regarded as exhaustive, which implies that even if prior odds are given, no conclusion can be drawn about the posterior probabilities of the hypotheses. We recommended that experts should abstain from reporting likelihood ratio's in such cases and should, more generally, explain to courts what are the implications of reporting on non-exhaustive hypothesis pairs.

Finally, although our analysis was confined to the reporting of probabilistic evidence by experts in trials, we believe it is also relevant for AI support for evidential reasoning. For instance, one of us (ARM) is leading a research project on teaching judges to analyse cases with the help of Bayesian network tools [8], which are an application of AI. One aim of this project is to test our above hypotheses on the possibly misleading effects of representation methods in experiments with human test subjects.

## Acknowledgement

## Declaration on generative AI

For the writing of this paper no generative AI was used.

## References

[1] P. Dawid, Probability and proof, 2005. Online appendix to T.J. Anderson, D.A. Schum and W.L. Twining: Analysis of Evidence, Boston, MA: Little, Brown and Company, 1991. URL: https://www.cambridge.org/us/download_file/203379/.

[2] R. Hill, Multiple sudden infant deaths – coincidence or beyond coincidence?, Pediatric and Perinatal Epidemiology 18 (2004) 320–326.

[3] W. Thompson, E. Schumann, Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and defense attorney's fallacy, Law and Human Behaviour 11 (1987) 167–187.

[4] C. Dahlman, A systematic account of probabilistic fallacies in legal fact-finding, The International Journal of Evidence and Proof 25 (2025) 45–64.

[5] J. Buckleton, D. Taylor, J.-A. Bright, T. Hicks, J. Curran, When evaluating DNA evidence within a likelihood ratio framework, should the propositions be exhaustive?, Forensic Science International: Genetics 50 (2021) 102406.

[6] R. Meester, M. Sjerps, Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence, Law, Probability and Risk 3 (2004) 51–62.

[7] H. Jellema, Reasonable doubt from unconceived alternatives, Erkenntnis 89 (2024) 971–996.

[8] A. Mackor, Risks of incorrect use of probabilities in court and what to do about them, in: A. Placani, S. Broadhead (Eds.), Risk and Responsibility in Context, Routledge, New York and London, 2024, pp. 94–108.