Chapter 5: Assessing evidential strength with the likelihood ratio

Rafal Urbaniak and Marcello Di Bello

The fallacies we considered in Chapter 3—base rate, prosecutor's and defense attorney's fallacy—show how the posterior probability of a hypothesis can be overestimated or underestimated. The posterior probability should reflect the evidence, but should not be identified with the probative value or strength of the evidence. A hypothesis may have a low posterior probability given the evidence, even though the evidence increases the probability of the hypothesis substantially.[1] So measuring evidential strength or probative value solely by posterior probabilities leaves out something crucial.

Another distinction worth making is between the global and local value of the evidence (Di Bello & Verheij, 2018). Let $E_1, E_2, \ldots, E_k$ be the total evidence presented at trial and H the ultimate hypothesis, say that the defendant is guilty of insider trading. The ultimate hypothesis is usually complex and can be thought of as the conjunction of several sub-hypotheses $H_1, H_2, \ldots H_k$. The total evidence bearing on the ultimate hypothesis should guide the final decision. But, as a preliminary step, it is useful to locally evaluate the impact of an individual piece of evidence E_i on the probability of a specific hypothesis H_i. When lay witnesses and experts testify at trial, the assessment of the evidential value of their individual testimonies should precede their aggregation into a whole, complex body of evidence.

This chapter articulates a probabilistic account of probative value or evidential strength that is incremental (it tracks changes in probability) and local (it is limited to individual pieces of evidence and specific hypotheses). We will argue that the likelihood ratio serves these purposes well. Section explains why it fares better than another popular measure of evidential strength, the Bayes factor. (Appendix **??** broadens the discussion to probabilistic measures of confirmation and reaches a similar conclusion.) We then offer two illustrations of how the likelihood ratio can be fruitfully deployed. Section shows that it allows for a nuanced assessment of the strength of quantitative evidence, DNA match evidence. Section **??** examines how it can help to evaluate eyewitness testimony This should dispel the impression that the likelihood ratio is only suited for explicitly quantitative evidence.

Despite its versatility, however, the likelihood ratio should be deployed with care. It can be hard to interpret in practice, as we discuss in Section **??**. In Section **??** we discuss evidential relevance, a topic closely related to that of evidential value. The likelihood ratio may categorize an item of evidence as irrelevant while intuitively the item is relevant. We explain this problem away by insisting that the likelihood ratio is a *local* measure whose meaning is relative

[1] Here is a more concrete example. Suppose an expert testifies that the blood found at the crime scene matches the defendant's and it is .05 probable that a person unrelated to the crime would match by coincidence. Absent other evidence to the contrary, it should initially be very likely that the defendant, as anyone else, had little to do with the crime. Say, for illustrative purposes, that the prior probability of the source hypothesis is .01, and let the probability of a match if the suspect is the source be approximately 1. By running Bayes' theorem, the posterior probability that the defendant is the source comes out to be roughly .17. While the match did not make it very likely that the defendant was the source of the traces, the posterior probability is seventeen times larger than the prior.

to specific hypotheses. To be sure, the value of an item of evidence is to be established both locally (relative to specific hypotheses) and globally (relative to the case as a whole). This suggests the need of formulating a more complex theory. We undertake this task in Part III.

The likelihood ratio is better than the Bayes factor

A popular measure of evidential strength is the *Bayes factor*, corresponding to the likelihood of the evidence—the probability of the evidence given the hypothesis of interest, $\mathrm{P}(E|H)$—divided by the probability of the evidence $\mathrm{P}(E)$:

$$\mathrm{BF}(E, H) = \frac{\mathrm{P}(E|H)}{\mathrm{P}(E)}.$$

It is a plausible measure as it appropriately deviates from one, its point of neutrality. Since, by Bayes' theorem, $\mathrm{P}(H|E)$ equals $\mathrm{BF}(H, E) \times \mathrm{P}(H)$, the Bayes factor is greater than one if and only if the posterior probability $\mathrm{P}(H|E)$ is higher than the prior probability $\mathrm{P}(H)$. The greater the Bayes factor (for values above one), the greater the upward shift from prior to posterior probability, the more strongly E positively supports H. Conversely, the smaller the Bayes factor (for values below one), the greater the downward shift from prior to posterior probability, the more strongly E negatively supports H. If $\mathrm{P}(H) = \mathrm{P}(H|E)$, the Bayes factor equals one and the evidence has no impact on H.

The posterior probability of H given E could still be low even when the Bayes factor is significantly above one, indicating that the evidence is strongly probative of H despite the low posterior probability. So, as desired, this measure captures a dimension of the value of evidence that is not reflected in the posterior probability. Unfortunately, the Bayes factor suffers from three shortcomings that make it unsuitable for applications in trial proceedings.

The first shortcoming is a dependency on the prior probability of the hypothesis of interest. To see why, consider the denominator $\mathrm{P}(E)$. It can be unpacked by the law of total probability:

$$\mathrm{P}(E) = \mathrm{P}(E|H)\mathrm{P}(H) + \mathrm{P}(E|\neg H)\mathrm{P}(\neg H). \tag{1}$$

So the Bayes factor can be written in a longer form:

$$\mathrm{BF}(E, H) = \frac{\mathrm{P}(E|H)}{\mathrm{P}(E|H)\mathrm{P}(H) + \mathrm{P}(E|\neg H)\mathrm{P}(\neg H)}. \tag{2}$$

What should be clear from this formulation is the dependency on the prior probabilities $\mathrm{P}(H)$ and $\mathrm{P}(\neg H)$. Indeed, suppose $\mathrm{P}(E|H) = 1$ and $\mathrm{P}(E|\neg H) = .1$. If $\mathrm{P}(H) = .1$, $\mathrm{P}(E)$, the denominator, is .19, and so the Bayes factor is approximately 5.26. If, however, $\mathrm{P}(H) = .2$, the denominator is $.28$ and the Bayes factor is approximately 3.57. In fact, a more general look (Figure **??**) shows that the prior probability can have larger impact on the Bayes factor than the likelihood $\mathrm{P}(E|\neg H)$.

This is a strike against adopting this measure of evidential strength in legal fact-finding. For suppose an expert who is testifying in court is tasked with assessing the value of an item of evidence, say a DNA or fingerprint match. This assessment should not depend on the expert's prior convictions about the plausibility of the hypothesis. Further, judges and lay jurors should be in a position to understand the expert's assessment in the same way, even if they assign different prior probabilities to the hypothesis.[2]

[2]The requirement of prior independence is also in line with an objectivity requirement that the strength of evidence should not vary from one researcher to another (Bickel, 2012).

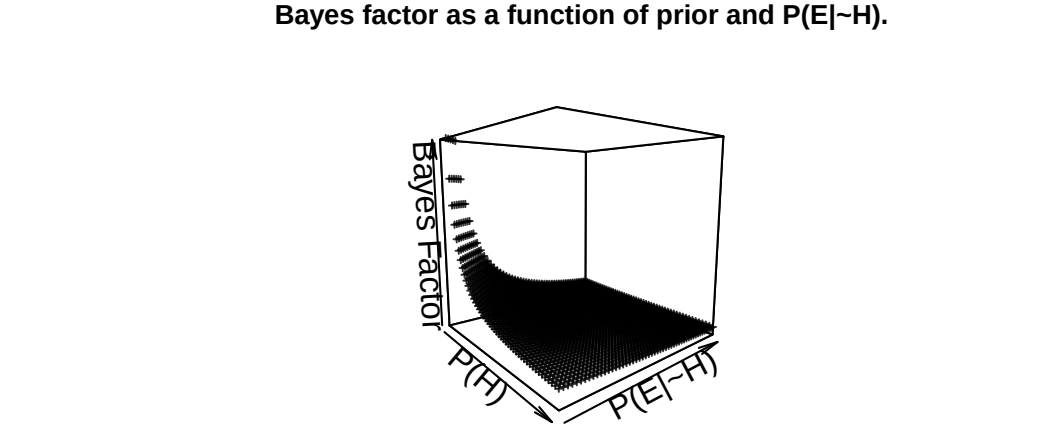

Figure 1: Impact of the prior and likelihood of E given ~H for probabilities in (0, 0.05) and Bayes Factor restricted to (0, 250) for visibility.

A second reason to worry about the Bayes factor is its complexity. To see this, the catch-all alternative hypothesis $\neg H$ in the denominator can be replaced by a more fine-grained set of alternatives, $H_1, H_2, \ldots H_k$, provided H and these alternatives are exclusive and cover the entire space of possibilities (that is, they form a partition). The denominator becomes:

$$\mathrm{P}(E) = \mathrm{P}(E|H)\mathrm{P}(H) + \sum_{i=1}^{k} \mathrm{P}(E|H_i)\mathrm{P}(H_i). \tag{3}$$

Assessing $\mathrm{P}(E)$ now looks quite difficult. It would require one to sift through the entire space of possibilities, as well as coming up with a sensible selection of priors probabilities for the several alternative hypotheses on hand. Whoever is tasked with assessing the strength of evidence—lay jurors, judges, or expert witnesses—might face too great a cognitive burden.

A third reason to hesitate about the Bayes factor comes from the problem of irrelevant conjuncts (Fitelson, 1999; Gillies, 1986). Consider the hypothesis H = 'the suspect is guilty of murder' and suppose it is a fact that E = 'the suspect killed the victim'. Fact E does not establish guilt with certainty since guilt requires both *actus reus*, the killing, and *mens rea*, the intention. But clearly E provides positive support for H. Now consider a composite hypothesis H' = 'the suspect is guilty of murder *and* we live in a simulation built by aliens'. Presumably, the support E provides for H' should be weaker than the support it provides for H. After all, the addition of a far-fetched hypothesis should weaken evidential support. This weakening, however, cannot be captured by the Bayes factor since both H and H' deductively entail E. In general, suppose $H \vDash E$ (and so, also, $H \wedge X \vDash E$). Then both $\mathrm{P}(E|H)$ and $\mathrm{P}(E|H \wedge X)$ equal 1. But this means that $\mathrm{BF}(H, E) = \mathrm{BF}(H \wedge X, E) = 1/\mathrm{P}(E)$. So, contrary to what one would expect, the Bayes factors for the two support relations are equal.[3]

[3]The same point can be made using irrelevant hypotheses that are not so far-fetched. For instance, suppose one hypothesis of interest is whether the victim was running in the park on a certain night, and the relevant piece of evidence is her footprints in the park. Perhaps, another hypothesis is whether she had wine at dinner later on.

The discussion so far suggests that a good measure of evidential strength should satisfy three desiderata: (i) it does not depend on priors, (ii) places no unreasonably heavy cognitive requirements, and (iii) does not fall prey to the problem of irrelevant conjuncts. We will argue that a measure that satisfies these desiderata is the *likelihood ratio*. It is defined as

$$\text{LR}(E, H, H') = \frac{\text{P}(E|H)}{\text{P}(E|H')},$$

where H' is a hypothesis that is a competing alternative to H. In the most straightforward case, H' is just the negation of H. But the competing hypotheses H and H' need not be one the negation of the other, a point to which we will return. As with the Bayes factor, support levels correspond to deviations from one. If the evidence is more likely given H than H', the ratio would be above one, and if the evidence is more likely given H' than H, the ratio would be below one. The greater the likelihood ratio (for values above one), the stronger the evidence in favor of H as contrasted with H'. The smaller the likelihood ratio (for values below one), the stronger the evidence in favor of the competing hypothesis H' as contrasted with H.

Note that both conditional probabilities in the numerator and denominator are needed for a correct assessment of the value of the evidence. For suppose we only relied on $\text{P}(E|H)$. In some cases, this conditional probability will be close to one. For instance, the probability that the blood from the crime matches the accused, if the accused is the source, $\text{P}(\text{blood match}|\text{source})$, may be close to one. Similarly, the probability that the DNA from the crime scene matches the accused, again if the accused is the source, $\text{P}(\text{DNA match}|\text{source})$, may also be close to one. Now, the DNA match should be stronger incriminating evidence than the blood type match because a specific genetic profile typically is less common than a specific blood type. Yet the quantity $\text{P}(E|H)$, by itself, makes no distinction here. The difference can instead be captured by the other conditional probability $\text{P}(E|\neg H)$. If the accused is *not* the source, the probability of a blood type match, while relatively small, should be higher than the probability of a DNA profile match. The likelihood ratio tracks both conditional probabilities, and thus it would be higher for the DNA match than the blood match, as desired.[4] For similar reasons, the strength of evidence cannot be measured by the probability $\text{P}(E|\neg H)$ alone. Consider an example by Triggs & Buckleton (2004). In a child abuse case, the prosecutor offers evidence that a couple's child rocks and that only 3% of non-abused children rock, $\text{P}(\text{child rocks}|\neg\text{abuse}) = .3$. If it is unlikely that a child who is not abused would rock, that this child rocks might seem evidence of abuse. But this interpretation is mistaken. It could also be that 3% of abused children rock, $\text{P}(\text{child rocks}|\text{abuse}) = .3$. If rocking is equally unlikely under either hypothesis, rocking cannot count as evidence of abuse.

So, both the probability of the evidence given the hypothesis and the probability of the evidence given an alternative hypothesis should be part of any good measure of evidential strength (ENFSI, 2015; Royall, 1997; Triggs & Buckleton, 2004). The Bayes factor includes both probabilities, but—as seen before—it falls prey to three difficulties. The likelihood ratio tracks both conditional probabilities without falling prey to these difficulties.

First, unlike the Bayes factor, the likelihood ratio does not depend on the prior probability of the hypothesis. This is is apparent from the odds version of Bayes' theorem:

$$\frac{\text{P}(H|E)}{\text{P}(H'|E)} = \frac{\text{P}(E|H)}{\text{P}(E|H')} \times \frac{\text{P}(H)}{\text{P}(H')}. \tag{4}$$

Clearly, whether she did is not obviously relevant to whether she was running in the park beforehand. However, one should be very hesitant to say that the evidential strength of the presence of footprints is the same relative to 'She was running in the park' and 'She was running in the park and had wine at dinner later on'. But this is what the Bayes factor would commit one to.

[4] Specifically, $\text{P}(\text{DNA match}|\text{source})/\text{P}(\text{DNA match}|\neg\text{source}) > \text{P}(\text{blood match}|\text{source})/\text{P}(\text{blood match}|\neg\text{source})$.

If the likelihood ratio is greater (lower) than one, the posterior odds will be greater (lower) than the prior odds of H. The likelihood ratio, then, is a measure of the upward or downward impact of the evidence on the prior odds of two hypotheses H and H'. This fits nicely with the division of labor common in legal fact-finding between experts and decision-makers, judges or lay jurors. A prominent forensic scientist recommends that 'in criminal adjudication, the values of the prior odds and the posterior odds are matters for the judge and jury ' (Colin Aitken & Taroni, 2008, p. 194). Other scholars recommend that experts should 'not trespass on the province of the jury …and should generally confine their testimony to presenting the likelihood of their evidence under competing propositions' (CGG Aitken, Roberts, & Jackson, 2010, p. 42). The meaning of the likelihood ratio can be made more perspicuous by supplementing it with a graph that conveys visually the extent to which the evidence changes the probability of the hypothesis of interest (See Figure **??**).[5]

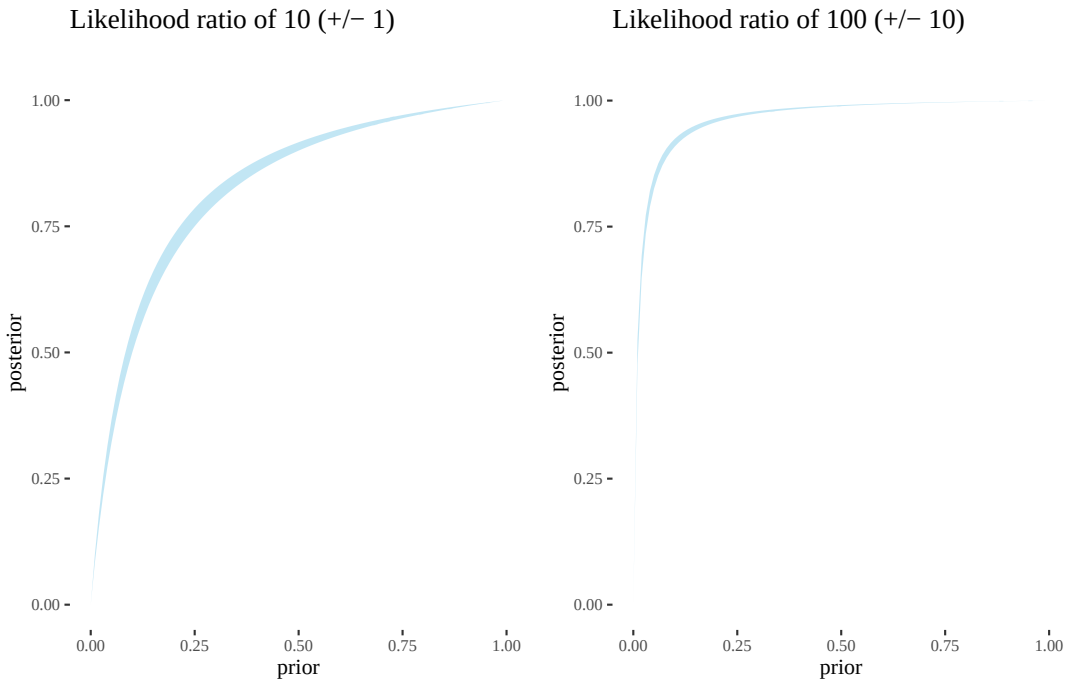

Figure 2: This graphical representation can supplement the likelihood ratio to convey visually the extent to which the evidence changes the probability of the hypothesis. This representation assumes that the two hypotheses in the likelihood ratio are one the negation of the other. In this case, the posterior probability of H equals $ ice-frac{PO}{1+PO}$, where PO are the posterior odds.

Second, the likelihood ratio is less cognitively burdensome than the Bayes factor. It does not require one to think about the probability of the evidence in general, P(E), say, the probability of a blood match under any possible scenarios. Direct and reliable estimation of this probability is difficult. From equation (1), it would require, besides an assessment of the conditional probabilities P$(E|H)$ and P$(E|\neg H)$, an assessment of the prior probabilities of P(H) and P$(\neg H)$. Instead, the likelihood ratio only requires an assessment of the conditional prob-

[5] Lund & Iyer (2017) have argued that the likelihood ratio often depends on the prior probabilities of the hypotheses under consideration. The reason is that the alternative hypothesis H' that occurs in the denominator can be analyzed as a disjunction of several alternative sub-hypotheses $H'_1 \vee H'_2 \vee \ldots \vee H'_k$. In this case, the denominator of the likelihood ratio would be the weighted sum of the likelihoods P$(E|H'_i)$, where the weights are the priors probabilities of each H'_i. We address this complication in later chapters after introducing the formal machinery of Bayesian networks.

abilities. In this sense, its calculation requires less information. This simplicity makes the likelihood ratio well-suited for presentation in trial proceedings. An expert, for instance, may testify that the blood-staining on the jacket of the defendant is ten times more likely to be seen if the wearer of the jacket hit the victim (prosecutor's hypothesis) rather than if he did not (defense's hypothesis) (CGG Aitken et al., 2010, p. 38).

Finally, unlike the Bayes factor, the likelihood ratio is not susceptible to the problem of irrelevant hypotheses. For suppose $\mathrm{P}(E|H) = \mathrm{P}(E|H \wedge X) = 1$, where X is an additional hypothesis that is irrelevant to H. Note that $\mathrm{LR}(E, H) = {}^{1}/_{\mathrm{P}(E|\neg H)}$, while $\mathrm{LR}(E, H \wedge X) = {}^{1}/_{\mathrm{P}(E|\neg H \vee \neg X)}$. Here the two denominators might differ. For example, suppose a fair coin is tossed three times. Let H = 'two first tosses resulted in two heads', E = 'at least one of the two first tosses resulted in a head', and X = 'the third toss resulted in heads'. Then $\mathrm{P}(E|H) = 1$, $\mathrm{P}(E|\neg H) = {}^{2}/_{3}$, $\mathrm{LR}(E, H) = \frac{1}{{}^{2}/_{3}} = 1.5$. However, $\mathrm{P}(E|H \wedge X) = 1$, $\mathrm{P}(E|\neg(H \wedge X)) \approx .71$, so $\mathrm{LR}(E, H \wedge X) \approx \frac{1}{.71} = 1.4$. Thus, the support, as measured by the likelihood ratio, can drop by adding a conjunct that is probabilistically irrelevant to the original hypothesis. In fact, this weakening of evidential support by adding an irrelevant conjunct holds in general for the likelihood ratio given sensible assumptions.[6]

All in all, the likelihood ratio outperforms the Bayes factor on several respects. But, of course, there could be other measures of evidential strength that fare even better. Other measures worth considering come from the literature in formal epistemology on confirmation theory. The expression 'confirmation' is more common in this literature than 'strength' (or value, support). A discussion of these measures, however, would detract us from the main task at hand. We therefore relegate it to Appendix **??**. In it, we show that the likelihood ratio is still, all things considered, the best measure on offer. A more general consideration to keep in mind is that there may well be two different questions here: (1) To what extend does a piece of evidence confirm our beliefs about a given hypothesis? (2) What is the strength (value, support) of a piece of evidence relative to a hypothesis? The two questions overlap to some extent. But the difference is that confirmation can depend on prior probabilities, while evidential strength should be kept separate from prior probabilities. Some confirmation measures may be seen as concerned with (1) rather than (2), and thus they are not always suitable for the evaluation of evidence in trial proceedings.

Match evidence and error probabilities

The two conditional probabilities that make up the likelihood ratio—for example, $\mathrm{P}(E|H)$ and $\mathrm{P}(E|\neg H)$—should be used in the evaluation of any form of evidence, both quantitative and non-quantitative. This section examines how a DNA match, a widely used form of quantitative evidence, should be evaluated by means of the likelihood ratio. The argument formulated here can be generalized to any 'match evidence'. The match can be between genetic profiles, fingerprints, blood types, bite marks, etc.

Consider an expert testimony that there is a genetic, DNA match between the traces at the

[6] Fitelson (2002) proved a general claim about irrelevant conjunctions. Hawthorne & Fitelson (2004) later strengthened this claim. The claim is that, if $\mathrm{LR}(E, H, \neg H) > 1$, $\mathrm{P}(E|X \wedge H) = \mathrm{P}(E|H)$, and $\mathrm{P}(X|H) \neq 1$, then $\mathrm{LR}(E, H, \neg H) > \mathrm{LR}(E, H \wedge X, \neg(H \wedge X))$. Crupi & Tentori (2010) raised a related problem. They point out that if $\mathrm{LR}(E, H) \leq 1$ and X is confirmationally irrelevant conjunct to H with regard to E, then E will have the same negative or null impact on $H \wedge X$, that is $\mathrm{LR}(E, H \wedge X) \leq \mathrm{LR}(E, H)$. They find this counter-intuitive and argue that this can be avoided by switching to the Z confirmation measure (Crupi, Tentori, & Gonzalez, 2007). As we argue in Appendix **??**, the Z measure is prior-sensitive and therefore not fit for our purpose. Further, the phenomenon might not be deeply troubling either. If the likelihood ratio tracks how strongly the evidence supports a hypothesis, it should be no surprise that a more complex hypothesis—one obtained by adding an irrelevant proposition—-enjoys a lower support from the same evidence.

crime scene and a sample from the defendant. This statement is evidence that the defendant was the *source* of the traces—that the materials found at the scene originated from the defendant. The match can also be evidence that the defendant was present at the scene or committed the crime, but these claims are more questionable, as the chain of inferences is weaker. For simplicity, let us focus on the source hypothesis. The likelihood ratio we should be concerned with is therefore the following:

$$\frac{\mathrm{P}(\mathrm{match}|\mathrm{source})}{\mathrm{P}(\mathrm{match}|\neg\mathrm{source})}$$

How strongly does a match favor the source hypothesis? When experts testify about a DNA match, they often only provide the so-called *random match probability* as an indicator of evidential strength. This quantity expresses the probability that a random person, unrelated to the crime, would coincidentally match the crime scene profile. The random match probability coincides (roughly) with the denominator of the likelihood ratio. But what about the numerator? The hidden assumption often made is that the numerator must be close to one. So the likelihood ratio is approximated by a simple formula:

$$\frac{\mathrm{P}(\mathrm{match}|\mathrm{source})}{\mathrm{P}(\mathrm{match}|\neg\mathrm{source})} \approx \frac{1}{\text{random match probability}}$$

Since the random match probability is usually an impressively low number, say 1 in 500 million, this is enough to ensure that the above ratio is significantly grater than one.

This analysis, though simple and elegant, lacks precision in at least two respects. First, it assumes that the numerator $\mathrm{P}(\mathrm{match}|\mathrm{source})$ is close to one. But a DNA match need not track with 100% probability the fact that the suspect is the source. There could be false negative matches. In addition—and more importantly—equating the denominator $\mathrm{P}(\mathrm{match}|\neg\mathrm{source})$ with the random match probability ignores the risk of false positive matches. This risk is not negligible (Shaer, 2016). The denominator, in fact, should depend on two sources of error: a false positive match and a coincidental match. These errors are quite distinct. For suppose two individuals—say the perpetrator and the defendant—happen to share the same DNA profile by coincidence. If an expert states that the crime scene sample and the defendant's sample match, this would be a coincidental match, not a false positive match. This risk of error is captured by the random match probability. But if the two samples do not actually match, and yet the expert says that they do, this would count as a false positive match, not a coincidental match. This risk of error is not captured by the random match probability.

Unlike a coincidental match, a false positive match is often caused by a human error in a number of circumstances (see (Thompson, 2013) for a more exhaustive treatment and multiple examples):

- **Cross-contamination of samples.** For instance, in Dwayne Johnson (2003) samples were accidentally swapped. In Lukis Anderson (2012), the genetic material was carried over by the paramedics. In one case, German police invested a considerable amount of time and effort searching for the so-called Phantom of Heilbronn, whose DNA profile was associated with many crimes. A bounty of EUR 300,000 was placed on her head. It turned out she was an innocent employee involved in the production of cotton swabs used across the country.
- **Mislabeling of samples.** For instance, in 2011 the Las Vegas Metropolitan Police Department acknowledged that samples of two men suspected of a 2001 robbery were switched, leading to the exclusion of the perpetrator and four years of incarceration for the other suspect. The mistake came to light only because the perpetrator was later arrested for another crime.

- **Misinterpretation of test results.** Single-source sample comparison is not easily prone to misrepresentation, but evidence mixtures—often needed in sexual assault cases—are complicated to interpret. For example, Dror & Hampikian (2011) re-examined a 2002 Georgia rape trial in which two forensic scientists had concluded that the defendant could not be excluded as a contributor of the crime traces. The evidence was sent to 17 lab technicians for re-examination. One of them agreed that the defendant could not be excluded as a contributor. Twelve considered the DNA exclusionary, and four found it inconclusive. If the quantity of DNA is limited, there is uncertainty about the number of contributors and about whether any alleles are missing. Ultimately, there is an element of subjectivity in mixed DNA interpretation.

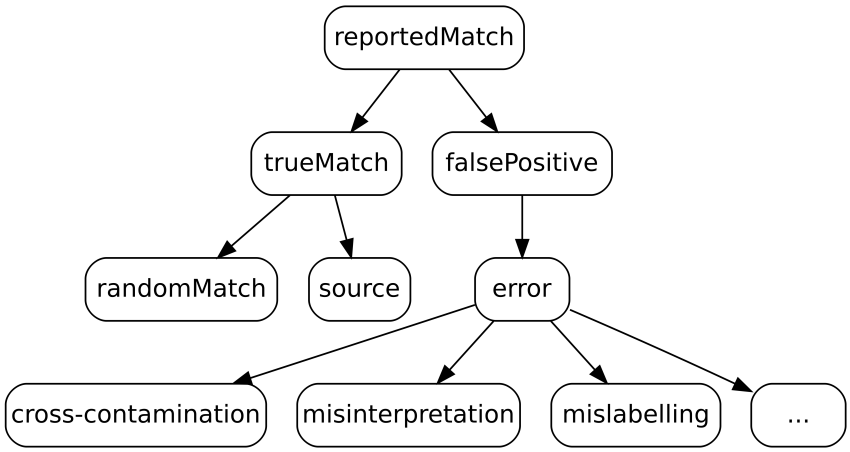

Figure 3: Dependencies between variables in the false positive problem.

The moral is that a careful evaluation of DNA evidence should take into account, besides the random match probability, the risks of false positive and false negative matches. This can be done using the likelihood ratio (**aitken2003probability?**). False positives are usually more worrisome than false negatives as they increase the risk of a mistaken conviction. That is also why we devoted to them more space in the foregoing discussion. But, for the sake of completeness, we will consider both.

This more refined analysis begins by making a conceptual distinction between true match and reported match. A true match is the fact that two samples actually carry the same genetic profile, while a reported match is a statement made by an expert that two samples match. A true match will exist not only if the suspect is the source, but also if, even though the suspect is not the source, the profiles are in fact the same due to a random, coincidental match. Similarly, a reported match might arise not only if there is a true match, but also if a false positive error has been made.[7] These possibilities are represented in Figure **??**. For ease of reference, we will use the following abbreviations:

[7]The notion of a reported match is a simplification. The expert's opinion may be more fine-grained. A reported match could be conclusive or merely probable. A non-reported match could be a definite exclusion or a prob-

S	The specimen comes from the suspect (source).
R	A match is reported (reported match).
M	There is a true match (true match).

In this set-up, the evidence to be assessed is the *reported* match relative to the pair of hypotheses S and $\neg S$. So the likelihood ratio we are after has the form:

$$\frac{\mathrm{P}(R|S)}{\mathrm{P}(R|\neg S)}.$$

With a few manipulations and assumptions in place, the likelihood ratio can be written as:[8]

$$\frac{\mathrm{P}(R|S)}{\mathrm{P}(R|\neg S)} = \frac{\mathrm{P}(R|M)}{\mathrm{P}(R|M)\mathrm{P}(M|\neg S) + \mathrm{P}(R|\neg M)\mathrm{P}(\neg M|\neg S)} \tag{5}$$

Note that, as intended, the numerator $\mathrm{P}(R|M)$ can be different from one, since a false negative reported match can occur (or, which is the same, a true match need not always occur). The denominator reflects the fact that there are two ways misleading evidence can arise: there is a true match and the suspect is not the source (because of a random, coincidental match), or there is no true match, and a false positive error has been made in the identification process.

To make the different sources of error more salient—false negative matches, false positive matches and random or coincidental matches—the likelihood ratio can be written, as follows:

$$\frac{\mathrm{P}(R|S)}{\mathrm{P}(R|\neg S)} = \frac{1 - FNP}{[(1 - FNP) \times RMP] + [FPP \times (1 - RMP)]} \tag{6}$$

This formula is the same as the earlier one. The expression FNP stands for the false negative probability $\mathrm{P}(\neg R|M)$, so $1 - FNP$ equals the true positive probability $\mathrm{P}(R|M)$. The expression FPP stands for the false positive probability $\mathrm{P}(R|\neg M)$. The expression RMP stands for the random match probability $\mathrm{P}(M|\neg S)$. A false positive or false negative probability track a human error, the possibility that a match may be reported (R) even without a true match ($\neg M$)

---

able exclusion. The expert could also testify that the laboratory analyses were inconclusive. So, instead of a binary report, match versus non-match, the expert could testify about a conclusive match, probable match, inconclusive laboratory results, probable exclusion, definite exclusion. This complexity would further complicate the analysis we present in this section, but would not invalidate the conceptual framework.

[8] By the law of total probability, the denominator $\mathrm{P}(R|\neg S)$ can be unpacked as $\mathrm{P}(R \wedge M|\neg S) + \mathrm{P}(R \wedge \neg M|\neg S)$. The latter, by the chain rule, is equivalent to $\mathrm{P}(R|M \wedge \neg S)\mathrm{P}(M|\neg S) + \mathrm{P}(R|\neg M \wedge \neg S)\mathrm{P}(\neg M|\neg S)$. A similar reasoning applies to the numerator. So we have:

$$\frac{\mathrm{P}(R|S)}{\mathrm{P}(R|\neg S)} = \frac{\mathrm{P}(R|M \wedge S)\mathrm{P}(M|S) + \mathrm{P}(R|\neg M \wedge S)\mathrm{P}(\neg M|S)}{\mathrm{P}(R|M \wedge \neg S)\mathrm{P}(M|\neg S) + \mathrm{P}(R|\neg M \wedge \neg S)\mathrm{P}(\neg M|\neg S)}$$

Both numerator and denominator can be simplified because a reported match (R), given a true match obtains (M), is independent of whether the suspect is the source (S):

$$\mathrm{P}(R|M \wedge S) = \mathrm{P}(R|M \wedge \neg S) = \mathrm{P}(R|M)$$

$$\mathrm{P}(R|\neg M \wedge S) = \mathrm{P}(R|\neg M \wedge \neg S) = \mathrm{P}(R|\neg M)$$

Finally, in the numerator, let the probability of a true match if the suspect is the source be one:

$$\mathrm{P}(M|S) = 1 \quad \text{so also} \quad \mathrm{P}(\neg M|S) = 0.$$

This assumption holds in virtue of the meaning of the statements involved. That the suspect is the source of the crime sample entails, almost analytically, that the two samples must carry the same genetic profile.

or that a match may *not* be reported ($\neg R$) even with a true match (M). The random match probability, instead, tracks a coincidence of nature, the possibility that someone who is not the source ($\neg S$) could still be—coincidentally—a true match (M). If we set FNP to 0, we obtain the same formula derived by (**aitken2003probability?**), who did not consider false negatives. Their simpler formula reads:

$$\frac{\mathrm{P}(R|S)}{\mathrm{P}(R|\neg S)} = \frac{1}{RMP + [FPP \times (1 - RMP)]}$$

Let's now examine the impact of the error probabilities FNP and FPP on the likelihood ratio, holding fixed certain values of the random match probability. Figure **??** shows the impact of error rates (for values between 0 and .05). Random match probabilities are assumed to be in the order of 10^{-9} (often reported in the case of two single-source samples over ten or more loci). A small increase in the false positive probability can lower the likelihood ratio dramatically. For instance, with FNP set at .05, the likelihood ratio drops from 10^8 to 19 as FPP goes from 0 to .05. Interestingly, however, the impact of the false negative probability FNP (for values between 0 and .05) is rather negligible. For instance, if FNP is .05, the likelihood ratio goes from 20 to 19 as FPP moves from 0 to .05.

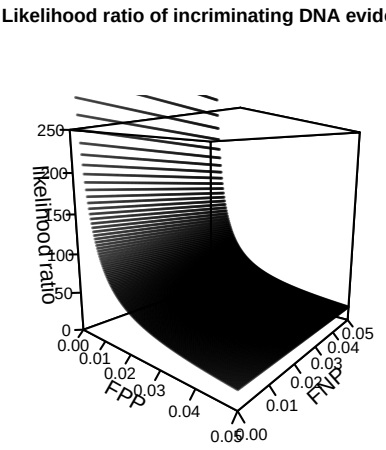

Figure 4: Impact of the error probabilities on the likelihood ratio of incriminatory DNA evidence for EQUATION (grid approximation). The impact of FNP is minor.

A similar analysis can be used to study the impact of error probabilities on the value of exculpatory DNA evidence, corresponding to a *negative* (reported) match $\neg R$. By replacing R with $\neg R$ in formula (5), the likelihood ratio becomes:

$$\frac{\mathrm{P}(\neg R|S)}{\mathrm{P}(\neg R|\neg S)} = \frac{\mathrm{P}(\neg R|M)}{\mathrm{P}(\neg R|M)\mathrm{P}(M|\neg S) + \mathrm{P}(\neg R|\neg M)\mathrm{P}(\neg M|\neg S)} \tag{7}$$

$$= \frac{FNP}{FNP \times RMP + [(1 - FPP) \times (1 - RMP)]} \tag{8}$$

Keep in mind that the negative reported match $\neg R$ is evidence *against* the source hypothesis S so long as the likelihood ratio is below one. At the extreme, if the false negative probability

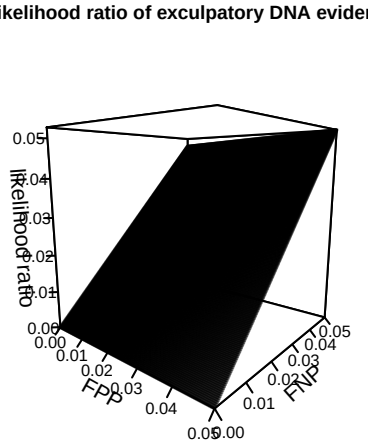

Figure 5: Impact of the error probabilities on the likelihood ratio of exculpatory DNA evidence, assuming EQUATION (grid approximation). The impact of FPP is minor.

FNP is zero, the numerator is zero. Thus, the likelihood ratio will be zero, as it should. In such a case, the negative match is completely exculpatory, and the posterior probability that the suspect is the source will also be zero. If the false negative probability is not zero, the greater the likelihood ratio (for values between 0 and 1), the weaker the value of the exculpatory match. As Figure **??** shows, the likelihood ratio progressively moves away from zero as the false negative error probability increases. For instance, with FPP fixed at 0.05, it goes from 0 to .052. Interestingly, however, the impact of the false *positive* probability FPP on the likelihood ratio of exculpatory evidence is essentially null. For instance, if FNP is fixed at .05, the likelihood ratio moves from .05 to .052 as FPP goes from 0 to .05.

So, we have seen that even a seemingly small error probability outweighs the random match probability. If there are good reasons to worry about random matches, there are even better reasons to worry about error probabilities. But the impact of the error probabilities is not uniform, contrary to what one might intuitively think. The false positive probability has a marked impact on the value of incriminating DNA evidence (positive matches), while the false negative probability has a marked impact on the value of exculpatory DNA evidence (negative matches). So this analysis shows which error probabilities we should be concerned about in which circumstances.

What actual numbers should we use for false positive and false negative probabilities? Unfortunately, no serious attempt has been made to systematically quantify the relevant error probabilities. Sometimes, a lab discovers its own errors and reports them, but this is rare (Thompson, 2013). Anecdotal information suggest that false positive matches take place more often than coincidental matches would entail, but how often remains unclear. Regular proficiency tests used in accredited DNA laboratories involve comparison of samples from known sources, but they are criticized for being unrealistically easy (yet, it happens that analysts fail them). Sometimes, corrective action files are made available. They usually show relatively

few false positive errors.[9] But because of the fragmentary data available, it is premature to conclude there is no reason for concern.

More data should be collected to plug in the right values of false positive and false negative probabilities in the likelihood ratio. Quite likely generic error frequencies will not be reliable enough, and relevant factors should be identified and used as predictors. To this end, error rate should be based on data that are fine-grained enough to document the error probabilities, FPP and FNP, corresponding to scenarios in which specific procedures, safeguards, or protocols are followed. As experts who testify about a match (or lack thereof) are cross-examined at trial, they could also testify about the procedures, safeguards and protocols that the laboratory technicians followed in the specific case. This case-specific information could then be used to estimate error probabilities under different scenarios. This would yield a more individualized assessment of the value of match evidence.

We conclude this section by noting that other proposals exist in the literature for formulating the likelihood ratio of a genetic match which also incorporate error probabilities. Another, more general proposal is due to Buckleton, Bright, & Taylor (2018). But, interestingly, the likelihood ratio of a DNA match used in equations (5) and (7) agrees with this other proposal given certain assumptions. This convergence is encouraging. Here we briefly go through the derivation.

First, Buckleton and co-authors make the conceptual distinction between the probability that an error occurs, $\mathrm{P}(\mathrm{ERR})$, and the probability that a match is reported if an error occurs, $\mathrm{P}(R|\mathrm{ERR})$. Let err denote the probability of error. The intuition here is that a laboratory error is an event in which the identification fails to be proper for technical or procedural reasons, for instance, the samples were switched, or some equipment failed and produced an erroneous reading. Whether such a laboratory error occurs should not depend on whether the source hypothesis holds. So the derivation starts with the assumption that err is probabilistically independent of the source hypothesis S:

$$\mathrm{err} = \mathrm{P}(\mathrm{ERR}) = \mathrm{P}(\mathrm{ERR}|S) = \mathrm{P}(\mathrm{ERR}|\neg S)$$

Separately, let k denote the probability of a reported match if an error occurs, also assumed to be independent of whether the source hypothesis is true:

$$k = \mathrm{P}(R|\mathrm{ERR}) = \mathrm{P}(R|\mathrm{ERR}, S) = \mathrm{P}(R|\mathrm{ERR}, \neg S)$$

Intuitively, even if normally a reported match R tracks to some extent the source of hypothesis S, say $\mathrm{P}(R|S) > \mathrm{P}(R|\neg S)$, this connection breaks down once an error err occurs. Now the derivation:

$$\begin{aligned} LR = \frac{\mathrm{P}(R|S)}{\mathrm{P}(R|\neg S)} &= \frac{\mathrm{P}(R|\neg\mathrm{ERR}, S)\mathrm{P}(\neg\mathrm{ERR}|S) + \mathrm{P}(R|\mathrm{ERR}, S)\mathrm{P}(\mathrm{ERR}|S)}{\mathrm{P}(R|\neg\mathrm{ERR}, \neg S)\mathrm{P}(\neg\mathrm{ERR}|\neg S) + \mathrm{P}(R|\mathrm{ERR}, \neg S)\mathrm{P}(\mathrm{ERR}|\neg S)} \\ &= \frac{1 \times (1 - \mathrm{err}) + k \times \mathrm{err}}{RMP \times (1 - \mathrm{err}) + k \times \mathrm{err}} = \frac{1 - \mathrm{err} + k \times \mathrm{err}}{RMP \times (1 - \mathrm{err}) + k \times \mathrm{err}} \end{aligned}$$

As before, the likelihood ratio is the ratio of the probabilities of a reported match if the suspect is the source and if the suspect is not the source. The numerator $\mathrm{P}(R|S)$ can be split into two possible scenarios: an error has not been made, or an error has been made. Accordingly, the numerator in the first line uses the law of total probability to split $\mathrm{P}(R|S)$ into these two

[9] For instance, the Santa Clara County district attorney's crime laboratory between 2003 and 2007 caught 14 instances of evidence cross-contamination with staff DNA, three of contamination by unknown person, and six of DNA contamination from other samples, three cases of DNA sample switch, one mistake in which the analyst reported an incorrect result, and three errors in the computation of the statistics to be reported.

scenarios. Similarly, the numerator $\mathrm{P}(R|\neg S)$ can be split into two cases: the suspect is not the source, but we are dealing with a random match, or the suspect is not the source, and an error has been made. An application of the law of total probability in the denominator mirrors this. The rest of the argument is just rewriting in terms of abbreviations, and algebraic manipulation.[10]

What is the connection of this formula to the one derived by (**aitken2003probability?**)? If an error guarantees a mistaken reported match, k becomes 1, err becomes the false positive rate, FPP. On this assumption, straightforward algebraic manipulation gives:

$$\begin{aligned}\frac{1 - FPP + 1 \times FPP}{RMP \times (1 - FPP) + 1 \times FPP} &= \frac{1}{RMP \times (1 - FPP) + FPP} \\ &= \frac{1}{RMP - FPP \times RMP + FPP}\end{aligned}$$

It takes a straightforward algebraic manipulation to show that this formula is identical to the one derived by (**aitken2003probability?**):

$$\frac{1}{RMP + FPP \times (1 - RMP)}$$

Aitken, CGG, Roberts, P., & Jackson, G. (2010). Fundamentals of probability and statistical evidence in criminal proceedings (Practitioner Guide No. 1), Guidance for judges, lawyers, forensic scientists and expert witnesses. *Royal Statistical Society's Working Group on Statistics and the Law*.

Aitken, Colin, & Taroni, F. (2008). Fundamentals of statistical evidence - a primer for legal professionals. *The International Journal of Evidence and Proof*, *12*(3), 181–207. https://doi.org/10.1350/ijep.2008.12.3.296

Bickel, D. R. (2012). The strength of statistical evidence for composite hypotheses: Inference to the best explanation. *Statistica Sinica*, 1147–1198.

Buckleton, J. S., Bright, J.-A., & Taylor, D. (2018). *Forensic DNA evidence interpretation*. CRC press.

Crupi, V., & Tentori, K. (2010). *Irrelevant conjunction: Statement and solution of a new paradox*. *77*(1), 1–13. https://doi.org/10.1086/650205

Crupi, V., Tentori, K., & Gonzalez, M. (2007). On Bayesian Measures of Evidential Support: Theoretical and Empirical Issues. *Philosophy of Science*, *74*(2), 229–252. https://doi.org/10.1086/520779

Di Bello, M., & Verheij, B. (2018). Evidential reasoning. In *Handbook of legal reasoning and argumentation* (pp. 447–493). Springer.

Dror, I. E., & Hampikian, G. (2011). Subjectivity and bias in forensic DNA mixture interpretation. *Science & Justice*, *51*(4), 204–208. https://doi.org/10.1016/j.scijus.2011.08.004

ENFSI. (2015). *Guidelines for evaluative reporting in forensic sciences*.

Fitelson, B. (1999). The plurality of bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science*, *66*, S362–S378. https://doi.org/10.1086/392738

Fitelson, B. (2002). Putting the irrelevance back into the problem of irrelevant conjunction. *Philosophy of Science*, *69*(4), 611–622. https://doi.org/10.1086/344624

Gillies, D. (1986). In defense of the popper-miller argument. *Philosophy of Science*, *53*(1), 110–113. https://doi.org/10.1086/289295

[10] The probability of a reported match R if no error occurs *and* the source hypothesis is false is the random match probability RMP, so $\mathrm{P}(R|\neg\mathrm{ERR}, \neg S) = RMP$. The probability that a reported match occurs when the source hypothesis is true and no error has made is assumed to be one, so $\mathrm{P}(R|S, \neg\mathrm{ERR}) = 1$.

Hawthorne, J., & Fitelson, B. (2004). *Discussion: Re-solving irrelevant conjunction with probabilistic independence*. *71*(4), 505–514. https://doi.org/10.1086/423626

Lund, S., & Iyer, H. (2017). Likelihood ratio as weight of forensic evidence: A closer look. *Journal of Research of National Institute of Standards and Technology*, *122*(27).

Royall, R. M. (1997). *Statistical evidence: A likelihood paradigm*. Chapman; Hall/CRC.

Shaer, M. (2016). The false promise of DNA testing. *The Atlantic*. Retrieved from https://www.theatlantic.com/magazine/archive/2016/06/a-reasonable-doubt/480747/

Thompson, W. C. (2013). Forensic DNA evidence: The myth of infallibility. In S. Krimsky & J. Gruber (Eds.), *Genetic explanations: Sense and nonsense* (pp. 227–347). Harvard University Press.

Triggs, C. M., & Buckleton, J. S. (2004). Comment on: Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence. *Law, Probability and Risk*, *3*, 73–82.