Probability and Statistics in the Law

Philip Dawid

Department of Statistical Science

University College London

Abstract

The field of legal reasoning is full of logical subtleties and probabilistic pitfalls. I survey a number of these, pointing out some of the problems and ambiguities, and various attempts to deal with them. Some celebrated court cases are used for illustration.

1 INTRODUCTION

Although the disciplines of Statistics and Law might seem far apart, they share some fundamental interests — in particular, the interpretation of evidence, testing of hypotheses, and decision-making under uncertainty. However, their differing backgrounds and approaches can often lead to misunderstandings, such as in the celebrated Collins case (Fairley and Mosteller 1977). The "New Evidence Scholarship" of the 1970's generated interest within some legal circles in the use of probability as an aid to rational interpretation of evidence, as well as some criticism. Some particularly interesting dialogues were stimulated by the pro-probability paper by Finkelstein and Fairley (1970) (see Tribe (1971b); Finkelstein and Fairley (1971); Tribe (1971a); Brilmayer and Kornhauser (1978); Kaye (1979)); and by the sceptical book by Cohen (1977) (see Schum (1979); Williams (1979); Cohen (1980); Williams (1980); Eggleston (1980)).

In recent years it has become apparent that problems arising in legal settings raise some fascinating and delicate issues of statistical logic, and that, in turn, proper application of statistical reasoning has a rôle to play in the pursuit of justice. In this paper I explore some of these logical issues, with reference to some real cases: see Dawid (2002) for some further background.

2 SUDDEN INFANT DEATHS

There have been a number of recent cases in the UK where two or more young children in a family have died suddenly from no obvious cause, and, even though there is no specifically incriminating evidence, their mother has been convicted of murdering them. In the case of Sally Clark, a paediatrician testified at trial that the probability P that her two sons would have died of SIDS (unexplained natural causes) was 1 in 73 million. That figure was widely and properly criticised, but it can not be denied that P is extremely small. The question is: What are we to make of such "statistical evidence"?

2.1 THE PROSECUTOR'S FALLACY

The correct interpretation $P = \Pr(\mathcal{E} \mid \overline{G})$ (where $\mathcal{E}$ denotes the evidence — here the fact of two infant deaths — G denotes "guilt" and $\overline{G}$ "innocence") is easily distorted into: $P = \Pr(\overline{G} \mid \mathcal{E})$. After all, to say that there is 1 chance in 73 million that the children died of natural causes appears to be just the same as saying that this is the probability that the mother did not kill them — seemingly overwhelming evidence for her being guilty. This mistaken "transposition of the conditional" is so common in court, where it usually favours the prosecution, that it has been termed "the prosecutor's fallacy" (see Gigerenzer (2002) for a clear account of the prosecutor's fallacy, and suggestions as to how it might be avoided). It would have been hard for Sally Clark's jury to ignore this seemingly powerful argument, and they did in fact convict.

2.2 COUNTER-ARGUMENT

There is an obvious counter-argument in this case, which I presented at appeal. We are comparing two alternative hypotheses: two deaths by SIDS, and two deaths by murder. If the chance of the former is relevant, should not that of the latter be equally relevant?

Using UK data, one could argue for a double murder figure of around 1 in 2 billion, to set against the SIDS figure of 1 in 73 million. One can see prosecution and defence brandishing their respective figures in adversarial combat, but the correct approach is to realise that it is their relative, not absolute, values that matter. In fact, their ratio $(1/2\text{ billion})/(1/73\text{ million}) = 0.0365$ can be interpreted as the odds on guilt given the evidence of the two deaths, implying a guilt probability of only 3.5%.

In the event, although the appeal court accepted that there had been some problems with the presentation of the statistical evidence at trial, it was not interested in properly identifying and understanding the logical issues involved. Sally Clark was eventually cleared on entirely unrelated grounds.

3 IDENTIFICATION EVIDENCE

Many criminal cases revolve about the issue of identity: is the suspect S the same person as the perpetrator C of the crime? Similar issues arise in civil cases, such as disputed paternity.

Forensic trace evidence is often brought in such cases. From the crime scene we obtain information I_C that can be assumed to apply to the criminal C — thus we may have a fingerprint, a footprint, fibres, or eyewitness evidence of sex, age, race, *etc.* With advances in DNA technology, it is now common to obtain a DNA profile of the criminal from biological material left at the scene of the crime. In addition, we have similar information I_S about the suspect S, for example his DNA profile. When this matches the crime sample, *i.e.* $I_S = I_C, = x$, say, that is clearly evidence in favour of the two samples having the same source. But how are we properly to weigh and apply this evidence?

One relevant feature of match evidence is the *match probability* P: this is the frequency with which the characteristic x occurs in the population at large. In the case of DNA profiling, the match probability can be estimated from population figures and genetic theory. Very tiny match probabilities, even as small as one in one billion, are now routine.

3.1 THE PROSECUTOR'S FALLACY

We henceforth implicitly condition on the suspect's characteristic: $I_S = x$. The match probability can be written as $P = \Pr(I_C = x \mid C \neq S)$. If we describe this as "the probability that the crime sample came from some one other than S", we are immediately in danger of committing the prosecutor's fallacy of § 2.1, which interprets P as $\Pr(C \neq S \mid I_C = x)$, *i.e.* the probability, in the light of the match, that S is in-

nocent — implying that the probability of guilt G is $1 - P$. If say $P = 0.0000001$, the jury or judge might well understand that the probability is only 1 in 10 million that S is not guilty, and convict.

3.2 THE DEFENCE ARGUMENT

A counter-argument along the lines of § 2.2 does not succeed here, since the probability of a match under the alternative hypothesis of guilt is unity.

Instead the defence might point out that there are $N + 1$ (say) people who could have committed this crime. One of these is truly guilty, and so matches the crime trace; while we would expect to see approximately NP innocent matches out of the remaining N innocent individual. We thus expect a total of $1 + NP$ matching individuals, of whom just 1 is guilty. If all we know about S is that he matches, the probability he is guilty is $1/(1 + NP)$. Taking $N = 30$ million and again $P = 0.0000001$, we would expect 3 innocent matches, for a final guilt probability of 1 in 4 — which is certainly not evidence "beyond a reasonable doubt".

3.3 SOME OTHER ARGUMENTS

The above defence argument can be varied in a number of ways (Dawid 1994), many of which are intuitively appealing — and have been recommended for use — but are in fact fallacious.

In all cases we assume that, prior to any evidence, any of the $N + 1$ individuals in the population is equally likely to be guilty, and that the only evidence $\mathcal{E}$ against S is that of the match: $I_S = I_C = x$. For illustration we take $N = 100$, $P = 0.004$.

Let M denote the unknown number of individuals i having $I_i = x$. We suppose that, before any samples are measured, M has the binomial distribution $\text{Bin}(N+1; P)$. We have $\Pr(G \mid \mathcal{E}, M) = M^{-1}$, and the final guilt probability, $\Pr(G \mid \mathcal{E})$, can be obtained by taking the expectation of this quantity with respect to the conditional distribution of M, given the evidence $\mathcal{E}$.

1. The evidence tells us that $M \geq 1$, and simple conditioning on this yields

$$\Pr(G \mid \mathcal{E}) = \mathrm{E}(M^{-1} \mid M \geq 1).$$

   For $M \sim \text{Bin}(N+1; P)$ this is not easily expressed in closed form, but can be calculated: for our numbers it evaluates to **0.902**.

2. An alternative argument is that, given the evidence, we know that there is one guilty match, and, out of the remaining N innocent individuals,

each has, independently, probability P of supplying a match. So the conditional distribution of M is $1 + \text{Bin}(N; P)$. Using this to take the expectation of M^{-1} yields

$$\Pr(G \mid \mathcal{E}) = \frac{1 - (1 - P)^{N+1}}{(N + 1)\, P}$$

which, for our values, gives **0.824**.

3. Finally, the correct approach.

   We can consider the total evidence ($I_C = x, I_S = x$) as the results, both successes, of two draws, *with replacement* (since C and S could be the same individual), from the population. The probability of this, given $M = m$, is $\{m/(N+1)\}^2$ and, using Bayes's Theorem, the resulting conditional distribution of M is

$$\Pr(M = m \mid I_C = x, I_S = x)$$
$$= \; cm \begin{pmatrix} N \\ m - 1 \end{pmatrix} P^{m-1} (1 - P)^{N-m+1}$$
$$(m = 1, \ldots, N + 1),$$

   where the normalising constant is $c = 1/(1+NP)$. Taking the expectation of M^{-1} with respect to this distribution then yields

$$\Pr(G \mid \mathcal{E}) = 1/(1 + NP),$$

   in agreement with the original (and much simpler) defence argument. This evaluates numerically to **0.714**.

The above is just one example of the pitfalls besetting logical and probabilistic reasoning in cases at law: see Balding and Donnelly (1995); Dawid and Mortera (1995); Dawid and Mortera (1996); Dawid and Mortera (1998) for a number of other subtle issues of interpretation of forensic identification evidence.

3.4   BAYES

A serious problem with both the prosecution and the defence arguments is that they do not allow for the incorporation of any other evidence in the case. The coherent approach to combining identification and other evidence is through Bayes's Theorem: *Posterior Odds* (on G) = *Prior Odds* × *Likelihood Ratio*, where the other evidence is accounted for in the prior odds, and the likelihood ratio based on evidence $\mathcal{E}$ (where here $\mathcal{E}$ is the match evidence "$I_C = I_S = x$") is defined by:

$$LR = \frac{\Pr(\mathcal{E} \mid G)}{\Pr(\mathcal{E} \mid \overline{G})}. \tag{1}$$

Because there is typically a subjective element in assessing prior probabilities, it is often argued that experts should confine their evidence to assessment of the more "objective" likelihood ratio, leaving the court to apply Bayes's Theorem with its own prior inputs. (However, see §§ 4 and 5 below concerning ambiguities in the definition of the likelihood ratio.)

In the case of identification evidence we can (usually) take $\Pr(\mathcal{E} \mid G) = 1$, $\Pr(\mathcal{E} \mid \overline{G}) = P$, so that the likelihood ratio is $1/P$. If the prior probability of guilt is π, the posterior probability is $\pi/(\pi + P - \pi P)$. This agrees (approximately) with the argument of the prosecutor when $\pi = 0.5$, and (exactly) with that of the defence when all $N + 1$ potential culprits are *a priori* equally likely to be the guilty party. This might be seen as support for the defence argument in the absence of any other evidence.

An interesting application of Bayes's Theorem was in the 1995 trial of Denis John Adams for sexual assault. The only prosecution evidence was a DNA match, with match probability assessed between 1 in 2 million and 1 in 200 million. The defence relied on the fact that the victim did not identify Adams at an identification parade, and also said that he did not look like the man who had raped her. In addition Adams's girlfriend testified that he had been with her at the time of the crime.

On the basis that the criminal was likely to be a local male aged between about 18–60, the prior probability of guilt, before any evidence, might be assessed at around one in 200,000. The likelihood ratio based on the DNA match is $1/P = 2$ million, say. That based on the victim's non-recognition of Adams could be assessed at, say, $0.1/0.9 = 1/9$, and that based on his girlfriend's alibi at, say, $0.25/0.5 = 1/2$. Assuming suitable independence, the posterior odds on guilt become $(1/200,000) \times (2,000,000) \times (1/9) \times (1/2) = 5/9$, corresponding to a posterior probability of 35% (though rising to 98% if we take $P = 1$ in 200 million).

In the actual case this argument was allowed at trial (although it does not seem to have impressed the jury, who convicted), but ruled out on appeal, on the basis that explaining how to think about probabilistic evidence "usurps the function of the jury", which "must apply its common sense". Unfortunately that leaves the door wide open to the prosecutor's fallacy and other tempting but misleading arguments.

4   DATABASE SEARCH

In some cases where a DNA profile is found at the crime scene there may be no obvious suspect. Then a trawl may be made through a police computer DNA

database in the hope that it will throw up a match. Suppose this happens: how, if at all, does the fact of the database search affect the strength of the evidence against a suspect so identified?

For definiteness, suppose that the database $\mathcal{D}$ is of size $n = 10,000$, that the match probability of the crime profile is $P = 1$ in 1 million, and that exactly one profile — that of S, say — in the database is found to match.

One intuition is that the database search has eliminated 9,999 individuals who would otherwise have remained alternative suspects. Given the very large initial number of alternative suspects, this has the effect of rendering the evidence in favour of S's guilt *very marginally stronger*. The relevant likelihood ratio is still close to 1 million.

An entirely different intuition proceeds by analogy with frequentist statistical approaches to testing multiple hypotheses. This would adjust the match probability to take account of the 10,000 possible ways of obtaining a match in the database, replacing it by the value, close to $10,000 \times (1$ in 1 million$) = 1/100$, of the probability of finding a match in the database, if it does not include the criminal. And a match probability of only 1 in 100 is *vastly weaker* evidence than one of 1 in 1 million. In particular, it corresponds to a likelihood ratio in favour of guilt of 100, rather than 1 million.

Whereas the former intuition focuses directly on the hypothesis H_S that S is guilty, the latter addresses this issue indirectly by focusing on the hypothesis $H_{\mathcal{D}}$ that some one in the database is guilty. Since we know there was exactly one match, to S, and only a matching individual can be guilty, these two hypotheses appear logically equivalent. Were this so, there would be no strong reason to focus on one rather than the other — which would be problematic, in view of the enormous difference between their associated likelihood ratios. Stockmarr (1999) has argued in favour of $H_{\mathcal{D}}$, and thus of quoting a likelihood ratio of 100, on the grounds that this hypothesis is data-independent, whereas hypothesis H_S can not even be specified in advance of performing the search and identifying S. However, while such data-dependence can affect frequentist inferences, its relevance to likelihood inference is arguable.

In fact, although hypotheses $H_{\mathcal{D}}$ and H_S are indeed equivalent once we know that the database $\mathcal{D}$ contains exactly one match, to S, they were not equivalent before making that observation: we may term them *conditionally equivalent*. A way of bridging the apparent chasm between them appears on realising that the *prior probability* of $H_{\mathcal{D}}$ is about 10,000 times larger than that of H_S. And, as we move between these hy-

potheses, this factor between their prior odds cancels exactly with that between their associated likelihood ratios noted above. Both approaches thus produce the identical posterior probability (whether for H_S or for $H_{\mathcal{D}}$ being unimportant, since these have truly become logically equivalent subsequent to the database search and its findings). If, taking a fundamental Bayesian position, we regard our inference as entirely carried by the posterior probability, there is thus no incompatibility between the two analyses.

We see that, in the absence of a clearly specified hypothesis, the concept of "the likelihood ratio" can not be regarded as objectively meaningful in itself, but rather is just one, volatile, ingredient of the (invariant) posteror inference — requiring the equally and oppositely volatile prior probability to complete that inference.

While this fully Bayesian analysis resolves the conceptual paradox, a serious practical problem remains. If "objectivity" requires that we offer likelihood ratios, rather than posterior probabilities, in evidence, which should we give? — and how can we ensure that its meaning and use is properly appreciated? As a matter of psychology it seems to me preferable to quote the likelihood ratio relating to H_S: the court will surely find it easier to understand and assess the prior probability that S is guilty, which is what is then needed to complete the analysis, rather than (as required for a $H_{\mathcal{D}}$-focused analysis) the prior probability that the guilty party is in the database. This choice is also legally preferable, since the fact that S was identified by searching a database may be inadmissible as evidence.

For further (often heated) discussion of these issues see Balding and Donnelly (1996); Donnelly and Friedman (1999); Stockmarr (1999); Evett *et al.* (2000); Dawid (2001).

5 MULTIPLE PERPETRATORS AND STAINS

A similar problem (Meester and Sjerps 2003; Meester and Sjerps 2004) arises when we know there were two criminals, two distinct DNA stains (say one on a pillow, one on a sheet) have been found at the scene of the crime, and there is a single suspect, S, who matches one of them — say the pillow stain — with its associated match probability P. How is the strength of the evidence against S affected by the multiplicity of stains?

Once again there is a choice of hypotheses to compare, these being logically equivalent in the light of the findings, but not in advance. A first approach compares

"S left one of the two stains" with "S did not leave either stain'; a second compares "S left the pillow stain" with "S did not leave either stain"; and yet a third compares "S left the pillow stain" with "S did not leave the pillow stain". Under some assumptions, the associated likelihood ratios are, respectively, $\frac{1}{2}P$, P, and $\frac{1}{2}P \times (2 - \delta)/(1 - \delta)$, where δ is the prior probability that S is guilty. And once again, the differences between these disappear after they are combined with their varying relevant prior odds. In Dawid (2004) I argue that it is the first of these likelihood ratios that relates most directly to the relevant issue: that of the guilt of S. But one must also take into account that the knowledge that there were two culprits effectively doubles the prior probability of S's guilt, as compared with a single-suspect case.

6   MIXED STAINS

In many cases, *e.g.* involving a rape or scuffle, a crime trace may clearly[1] be a mixture of biological material from more than one individual. We may or may not know how many contributors are involved, or the identity of some of them. It is sometimes possible to separate out the components of different contributors, *e.g.* by taking into account the differing amounts of DNA at different bands, but this is unreliable.

Suppose we have a suspect S who "matches" the crime trace, in that all his bands are contained in it. What is the strength of the DNA evidence against him? This can involve complex and subtle calculations and be sensitive to assumptions made.

6.1   O. J. SIMPSON

In the celebrated trial of O. J. Simpson for double murder, one of the crime samples could be explained as a mixture of blood from Simpson and one of the victims, Ron Goldman. At a certain locus, Simpson had genotype AB, Goldman AC, and the crime sample had ABC. In pre-trial depositions[2], the prosecution argued that the relevant match probability P should be taken as the frequency of Simpson's genotype AB — about 5%. (Such a P would be multiplied by similar figures calculated for other loci to obtain an overall match probability). The defence argued that P should be the total probability of any of the genotypes, AA, AB, AC, BB, BC, CC, that would have "matched" the crime sample: about 39%.

However, on the assumption that the mixture consists of Goldman and the culprit, the culprit must have type

---

[1] For example, because it has more than two bands at some locus.

[2] http://tinyurl.com/2fhsx

AB, BB or BC. These have combined probability 21%, and it is the reciprocal of this figure for P that yields the correct likelihood ratio. If we did not know Goldman's genotype, or thought that the other contributor was some one else, we need to conduct a more complex calculation to obtain the relevant likelihood ratio. Interpreting this as P^{-1}, we again obtain $P \approx 21\%$ (though this is an accidental concurrence of two potentially different figures).

7   MISSING SUSPECT

When a suspect, or other relevant party, is not available for DNA profiling, useful information can sometimes be obtained by profiling relatives — although the analysis then required can be both conceptually and computationally challenging.

7.1   HANRATTY

In 1962 James Hanratty was executed for rape and murder. In 1998 a DNA profile, assumed to be from the culprit, was extracted from some items that had been stored since the crime. Its associated match probability was around 1 in 2.5 million. Ever ready to fall for the prosecutor's fallacy, the Press duly reported this as "There is a 1 in 2.5 million chance that Hanratty was not the A6 killer" — even though, since Hanratty's DNA was unavailable, there was no more evidence against him than against any one else.

Hanratty's mother and brother now offered their own DNA for profiling — and this failed to exclude him. Again reports of the above match probability circulated as evidence of his guilt. In fact, the actual likelihood ratio, based on the indirect evidence of his relatives' DNA, was around 440.

Finally his body was exhumed, and a direct match obtained. Although the defence attempted to attribute this to contamination, it is generally agreed that the case is now closed.

7.2   DISPUTED PATERNITY

Problems of disputed paternity necessarily revolve around indirect "matching" of the DNA of the putative father with that of the true father, as partially revealed through the child's DNA. When profiles from mother, child and putative father are available, the likelihood ratio in favour of paternity can be calculated by standard formulae. When the putative father's profile is unavailable, profiles may be obtained from his relatives: for example, two full brothers, and an undisputed child and its (different) mother. Although the logical steps in calculating the likelihood

ratio are clear in principle (though not always so to the forensic and other experts directly involved in such work), the computational difficulties of implementing them can be severe.

8 BAYESIAN NETWORKS

In many court cases there is a mixed mass of evidence, the various items relating to each other and to the ultimate issue in complex and subtle ways. Whereas most lawyers are content to handle these complexities in purely intuitive ways, some legal scholars have considered formal tools to help in this process. An early and still influential contribution was the development of the *Wigmore chart* (Wigmore 1913; Wigmore 1937), a graphical representation of qualitative relationships between items of evidence.

There is some *prima facie* similarity between Wigmore charts and the modern technology of *Bayesian networks* (Cowell *et al.* 1999). Bayesian networks have been used for both qualitative and quantitative analysis of legal evidence. The former was undertaken by Dawid and Evett (1997) in the context of a prosecution for robbery, where it was required to combine eyewitness, fibre and blood evidence: the structure of the network implies various conditional independence relationships between the variables, that can be extracted and used to simplify expressions for likelihood ratios. Another fascinating use of Bayesian networks (which was also strongly influenced by Wigmorean analysis) is the reanalysis by Kadane and Schum (1996) of the evidence in the celebrated murder trial of Sacco and Vanzetti.

Bayesian networks have proved particularly valuable in addressing complex problems of interpretation of DNA profile evidence. Thus Figure 1 shows a Bayesian network representation of the absent father paternity case described in §7.2. This displays evidential relationships in a semantically unambiguous, clear and striking manner, and supports complex computational analysis using general purpose Probabilistic Expert System software such as Hugin[3] — see Dawid *et al.* (2002) for further details. This technology is also being applied to still more complex problems of DNA analysis, involving, separately or in combination, such features as mixed stains (Mortera *et al.* 2003), mutation (Dawid *et al.* 2001; Dawid 2003), contamination, and field and laboratory errors.
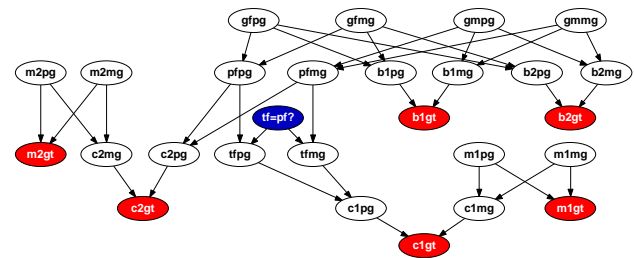
---
[3]http://www.hugin.com

Figure 1: Bayesian Network Representation of a Complex Paternity Case

9 CONCLUSIONS

Seemingly straightforward problems of legal reasoning can quickly lead to complexity, controversy and confusion: the above examples are just a few amongst many. The field provides a rich and challenging testbed for any general approach or technique for reasoning under uncertainty. Success in this venture could also have real impact on the fairer administration of justice.

References

Balding, D. J. and Donnelly, P. J. (1995). Inference in forensic identification (with Discussion). *Journal of the Royal Statistical Society, Series A*, **158**, 21–53.

Balding, D. J. and Donnelly, P. J. (1996). DNA profile evidence when the suspect is identified through a database search. *Journal of Forensic Sciences*, **41**, 603–7.

Brilmayer, L. and Kornhauser, L. (1978). Quantitative methods and legal decisions. *University of Chicago Law Review*, **46**, 116–53.

Cohen, L. J. (1977). *The Probable and the Provable*. Oxford : Clarendon Press.

Cohen, L. J. (1980). The logic of proof. *Criminal Law Review*, **1980**, 91–103.

Cowell, R. G., Dawid, A. P., Lauritzen, S. L., and Spiegelhalter, D. J. (1999). *Probabilistic Networks and Expert Systems*. Springer, New York.

Dawid, A. P. (1994). The island problem: Coherent use of identification evidence. In *Aspects of Uncertainty: A Tribute to D. V. Lindley*, (ed. P. R. Freeman and A. F. M. Smith), chapter 11, pp. 159–70. John Wiley and Sons, Chichester.

Dawid, A. P. (2001). Comment on Stockmarr's "Likelihood ratios for evaluating DNA evidence, when the suspect is found through a database search" (with response by Stockmarr). *Biometrics*, **57**, 976–80.

Dawid, A. P. (2002). Bayes's theorem and weighing evidence by juries. *Proceedings of the British Academy*, **113**, 71–90. (Published in book form as *Bayes's Theorem*, edited by Richard Swinburne).

Dawid, A. P. (2003). An object-oriented Bayesian network for estimating mutation rates. In *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, Jan 3–6 2003, Key West, Florida,* (ed. C. M. Bishop and B. J. Frey). http://tinyurl.com/39bmh.

Dawid, A. P. (2004). Which likelihood ratio? In discussion of Meester and Sjerps (2004). *Law, Probability and Risk*, **3**, 65–71.

Dawid, A. P. and Evett, I. W. (1997). Using a graphical method to assist the evaluation of complicated patterns of evidence. *Journal of Forensic Sciences*, **42**, 226–31.

Dawid, A. P. and Mortera, J. (1995). In discussion of Balding and Donnelly (1995). *Journal of the Royal Statistical Society, Series A*, **158**, 46.

Dawid, A. P. and Mortera, J. (1996). Coherent analysis of forensic identification evidence. *Journal of the Royal Statistical Society, Series B*, **58**, 425–43.

Dawid, A. P. and Mortera, J. (1998). Forensic identification with imperfect evidence. *Biometrika*, **85**, 835–49.

Dawid, A. P., Mortera, J., and Pascali, V. L. (2001). Non-fatherhood or mutation? A probabilistic approach to parental exclusion in paternity testing. *Forensic Science International*, **124**, 55–61.

Dawid, A. P., Mortera, J., Pascali, V. L., and van Boxel, D. W. (2002). Probabilistic expert systems for forensic inference from genetic markers. *Scandinavian Journal of Statistics*, **29**, 577–95.

Donnelly, P. J. and Friedman, R. D. (1999). DNA database searches and the legal consumption of scientific evidence. *Michigan Law Review*, **97**, 931–84.

Eggleston, R. (1980). The probability debate. *Criminal Law Review*, **1980**, 678–88.

Evett, I. W., Foreman, L. A., and Weir, B. S. (2000). Letter to the Editor (with responses by A. Stockmarr and B. Devlin). *Biometrics*, **56**, 1274–5.

Fairley, W. B. and Mosteller, F. (1977). A conversation about Collins. In *Statistics in Public Policy*, (ed. W. B. Fairley and F. Mosteller), pp. 369–79. Addison-Wesley, Reading, Massachusetts.

Finkelstein, M. O. and Fairley, W. B. (1970). A Bayesian approach to identification evidence. *Harvard Law Review*, **83**, 489–517.

Finkelstein, M. O. and Fairley, W. B. (1971). A comment on "Trial by Mathematics". *Harvard Law Review*, **84**, 1801–9.

Gigerenzer, G. (2002). *Reckoning with Risk: Learning to Live with Uncertainty*. Allen Lane: The Penguin Press.

Kadane, J. B. and Schum, D. A. (1996). *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*. John Wiley and Sons, New York.

Kaye, D. (1979). The laws of probability and the law of the land. *University of Chicago Law Review*, **47**, 34–56.

Meester, R. W. J. and Sjerps, M. (2003). The evidential value in the DNA database search controversy and the two stain problem. *Biometrics*, **59**, 727–32.

Meester, R. W. J. and Sjerps, M. (2004). Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence (with discusion by A. P. Dawid, D. J. Balding, J. S. Buckleton and C. M. Triggs). *Law, Probability and Risk*, **3**, 51–86.

Mortera, J., Dawid, A. P., and Lauritzen, S. L. (2003). Probabilistic expert systems for DNA mixture profiling. *Theoretical Population Biology*, **63**, 191–205.

Schum, D. (1979). A review of the case against Blaise Pascal and his heirs. *Michigan Law Review*, **77**, 446–83.

Stockmarr, A. (1999). Likelihood ratios for evaluating DNA evidence when the suspect is found through a database search. *Biometrics*, **55**, 671–7.

Tribe, L. H. (1971a). A further critique of mathematical proof. *Harvard Law Review*, **84**, 1810–20.

Tribe, L. H. (1971b). Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review*, **84**, 1329–93.

Wigmore, J. H. (1913). The problem of proof. *Illinois Law Review*, **8**, 77–99.

Wigmore, J. H. (1937). *The Science of Judicial Proof*, (third edn). Little, Brown, Boston.

Williams, G. (1979). The mathematics of proof. *Criminal Law Review*, **1979**, 297–308 and 340–54.

Williams, G. (1980). A short rejoinder. *Criminal Law Review*, **1980**, 103–7.