# THE EXPERT's EVIDENCE AND THE JUDGE's EVALUATION OF ALL INFORMATION IN COURT DECISIONS

Hummel K, Institute for Blood-Group Serology and Genetics
POB 880 D-79008 Freiburg, Germany

A great part of legal judgments rests on deterministic evidence. Such evidence is always sufficient in law. But evidence may also be of a circumstantial, ie, a "probabilistic" nature. This type of evidence is not as "absolute" as the former. In consequence, an important role is reserved to the judge's subjective conviction, ie: he reaches his conclusion through the sum of the circumstances.

In reaching a decision the judge holds the sum of evidence to be either "sufficient in law" or dismisses it as insufficient. An Appeal Court may reach a different conclusion - which, given the nature of probabilistic evidence - should not surprise. This does not imply arbitrariness in matters of judgment. It is a reflection of the fact that each individual judge makes his own independent evaluation, which, in most cases, is not - indeed, cannot be - a purely formal matter; the assessment of evidence is 'ipso facto' creative.

Litigation makes sense only if allegations appear to be based on facts. In this context, allegation should have some initial probability, ie, some evidence for the initial suspicion. Without facts of this nature it would be impossible to institute any legal proceedings. Hence, all legal proceedings start with a foreknowledge. In most cases these is fairly limited. The judge therefore requires additional informations to reach a legal decision. The additional evidence is "added" to the initial probability. If the latter is large, less additional evidence is needed than if it were small. When experts provide facts to help the judge to reach a decision these are often in the form of numerical values that support either the null or the counter hypothesis. As mentioned above, the experts must be satisfied that the two alternatives are realistic, in other words, that they have prior probabilities. For this reason, in his formula Bayes (2) allotted a "prior probability" to the frequencies of

Advances in Forensic Haemogenetics 5 (c) Springer-Verlag Berlin Heidelberg

each hypothesis. In accordance with his "postulate", one uses a neutral one, ie. the same prior probability for each of the hypotheses (unless specific numerical values are available).
A balance may help to illustrate the situation. One arm represents the null hypothesis, the other one the counter hypothesis. Initially, the pans, which will hold the informations presented in the case, are in equilibrium, as shown by the vertical position of the needle. One assumes "a priori" that both hypotheses have an equal chance of being true, ie, have the same initial probability. Using data he has obtained himself or has at his disposal the expert works out the weights that each hypothesis should have and puts them in the appropriate pans. The judge adds further (usually non-numerical) arguments from the rest of the evidence. He then studies the final position of the needle and tries to draw a convincing conclusion.
Let us term the null hypothesis X and the counter hypothesis Y. The expert obtains a weight in favour of X = G(X) and a different one in favour of Y = G(Y).
The ratio between them, L = G(X)/G(Y), is termed the likelihood ratio. By standardizing the values for G(X) and G(Y), one obtains the probabilities $W_X$ and $W_Y$ for both hypotheses:

$$W_X = G(X)/\left[G(X)+G(Y)\right] ;$$
$$W_Y = G(Y)/\left[G(X)+G(Y)\right] = 1-W_X .$$

Bayes was the first to correctly apply this formula to the individual case. De Laplace(4), Boole(3), Ball(1) and v.Mises(5) subsequently treated the problem analytically. Ever since, Bayes' Theorem (as it is generally known) not only has great legal significance, but also plays an equally important role in medical diagnostics, industrial production, archaeological research, etc.
Bayes, an Englishman, was a Presbyterian minister. His preoccupation with mathematics was a private interest. The treatise on which his fame rests was found in a drawer of his bedside table after his death in 1761. The mathematician Price recognized the importance of the work and shortly afterwards presented it in a lecture at the Royal Society of London.
One and a half centuries after Bayes, in the mid 1930s, another method of decision procedure took its place besides Bayes' Theorem: the Neyman-Pearson model(6). It works with limits and

Advances in Forensic Haemogenetics 5 (c) Springer-Verlag Berlin Heidelberg

type I and type II errors. Bayes' and Neyman-Pearson's methods have their respective fields of application. In terms of the court setting one may describe Bayes'as that of the expert and Neyman-Pearson's as that of the judge.

Of the two, the Neyman-Pearson model is much the easier to comprehend. Misunderstandings and a lack of appreciation of Bayes' principle have led to unjustified criticism and a strong anti-Bayes' body of opinion, this for a long time. The misunderstandings arise from the process of transferring the weighted ratios of mass statistics to the individual case. As Ball and v.Mises put it: there are no statistics for individuals and no individuals for statistics. That is true. But one can treat an individual - or a single event - in relation to a reference class, ie, a collective, and allocate a statistical probability to this person or event as a member of a certain reference class. Two reference classes correlate with the hypothesis pair, eg. father - non-father, instigator - non-instigator, perpetrator - non-perpetrator. To define the two reference classes one employs as many pertinent factors as possible, eg, in paternity cases a large number of hereditary characteristics in order to distinguish fathers from non-fathers. The greater the class-typing information, the greater the certainty of allocating a specific individual or case to one or other reference class. Of course, it is always true that any single event is unique and cannot, therefore, be repeated. Consequently, a single event cannot, strictly speaking, belong to a certain reference class.

In addition to the objection that every probability statement is relatively indeterminate with respect to the individual case, there is the objection that the empirical values for very narrowly defined reference classes are not fully reliable in terms of mass statistics. Lothar Sachs(7) puts the objections in perspective elegantly and pragmatically: reliable statements are imprecise, precise statements are unreliable. Any expert who knows his biostatistics is aware of this problem. He takes adequate account of it by applying statistics to the individual case only if this case can be allocated to a sufficiently narrow and, at the same time, sufficiently large reference class.

Only the judge has all the evidence pertaining to a case at his

Advances in Forensic Haemogenetics 5 (c) Springer-Verlag Berlin Heidelberg

disposal. The experts charged with collecting material evidence as well as the persons responsible for examinations, interrogations and investigations all provide pieces of evidence that alone are not, as a rule, decisive for the verdict. In most cases, no single individual providing these auxiliary services will have heard or seen all the evidence.

All the evidence is gathered together for the first time in the judge's summing-up. He weighs up the pros and cons of each individual piece of evidence, assessing each piece in relation to the rest as well as to a specific person, either the defendant or the plaintiff. This process produces a "probability of incrimination". Not all individual facts refer directly to people involved in the case. For instance, if traces of paint or tools have been identified, it is the duty of the judge to relate these findings to a suspect; his conclusion may be either positive or negative. In other instances the evidence may refer directly to the defendant, eg, a high probability of paternity or an "identical" blood test. But even in these apparently relatively uncomplicated instances the judge must associate the findings with a person; for this alone enables him to pass judgment.

Thus, the purpose of gathering evidence and relating the individual pieces to each other is to characterize and individualize the case as far as possible. This enables us to establish its type, and allocate it to as small a group of analogous cases as possible. If this particular group is also part of the judge's body of experience, the resulting judgment will fit the facts. The more strongly the case is focussed and the more differentiated the judge's body of experience, the sooner judgement will be passed and the more just it will be.

After the weighted evidence as a whole has been considered the judge has two bodies of arguments at his disposal: one in support of the allegation and the other against the allegation. He must now decide whether the probability of the allegation (a purely conceptual value) exceeds a - similarly conceptual - assessment limit or not. The judge himself determines the assessment limit. He uses principles of "utility", in that he tries to achieve a balance between the interests of the respective parties - or of the accused and the general public, all in the

Advances in Forensic Haemogenetics 5 (c) Springer-Verlag Berlin Heidelberg

light of the evidence presented in the actual case. He will set the assessment limit in a way that optimizes the benefit he seeks to provide through his judgment. To better picture the degree of the benefit associated with the decision limit, he will privately orient himself by similar cases in his own body of legal experience and cases in the literature as well as by his own experience of life in general. If the value for the present case (either pro or contra the "null hypothesis") lies above this decision limit he will be able to pronounce judgment: the action will be either upheld or dismissed. If the value lies below this decision limit, it will be necessary to seek further evidence.

The statistician will have noticed that the underlying principle of this process is not Bayes'but another. There is a "posterior probability" in Bayes'sense, ie, an integration of all evidence, including a prior proportion. This probability value is not assessed as such. It is incorporated into the distribution of values of other cases within the judge's body of experience, in which actions upheld and actions dimissed have their own respective distributions. This procedure corresponds to the Neyman-Pearson decision model.

A significant factor is the quota of error to be expected eg for non-fathers falsely ruled to be fathers, ie, the tail probability for non-fathers at the point applicable to the specific case. For the most cases of disputed paternity it is very low, practically zero. A ruling is thus possible.

Why does the judge use Neyman-Person instead of Bayes? The answer is simple. No judge will have such breadth of experience as to have dealt with 100 cases with the same circumstances. In reality, in no two cases are the circumstances exactly the same. Thus, it is impossible for the judge to decide according to Bayes, to declare that on the given evidence one may expect for instance that 99 of 100 cases with the same circumstances will be decided correctly and one falsely. Rather, the judge must picture the present case as one of a series of comparable (even if not identical) cases so as to decide for himself whether it better fits the distribution of upheld or the distribution of the dismissed cases. In other words, he must use the Neyman-Pearson distributions to orient himself about the statistically

Advances in Forensic Haemogenetics 5 (c) Springer-Verlag Berlin Heidelberg

error when - for the given decision limit - the case is decided one way or the other.
If judges were to orient themselves by Bayes'expectation of error, the level of their demands upon the evidence would always be the same, independent of, say, progress in medico-scientific opinions. For example, the demands placed upon serostatistical evidence of real paternity would remain at eg W = 95 %. In reality, over the years judges have demanded increasingly higher probabilities for their rulings. Therefore, they obviously orient themselves by the actual level of W values produced in similar cases. By choosing the Neyman-Pearson model instead of Bayes' principle, they demonstrate their awareness of the contemporary potential of evidence and, thereby, their flexibility and acceptance of progress.

## REFERENCES

1. Ball VC (1961) The moment of truth probability theory and standards of proof. In: Essays on Procedure and Evidence; edit.: Roaly and Covington; Nashville
2. Bayes Th (1763) An essay towards solving a problem in the Doctrine of Chances; communicated by R.Price, Philos.Transact.Roy.Soc. London 53, p.370-375
3. Boole G (1857) On the application of the theory of probabilities to the question of the combination of testimonies or judgments; Keith-Price Essay. In: Studies in Logic and Probability, London 1952, p.308-385
4. Laplace PS, de (1812) Théorie analytique de probabilité. 3me éd. Paris 1920
5. Mises R, von (1936) Wahrscheinlichkeit, Statistik und Wahrheit, 3.Aufl. Wien 1951
6. Neyman J, Pearson ES (1935), in: Kendall MG (edit.) The Advanced Theory of Statistics 2, 3rd ed. London 1951
7. Sachs L (1978) Angewandte Statistik (Statistische Methoden und ihre Anwendung) 5.Aufl. Berlin

Advances in Forensic Haemogenetics 5 (c) Springer-Verlag Berlin Heidelberg