RISK ANALYSIS AND JURISPRUDENCE; A RECENT EXAMPLE

Roger Cooke
Department of Mathematics
Delft University of Technology
Delft, The Netherlands
r.m.cooke@its.tudelft.nl

Cor Kraaikamp
Department of Mathematics
Delft University of Technology
Delft, The Netherlands

ABSTRACT

We discuss a recent incident in which risk analysts appeared as expert witnesses in a civil tort case. The problems which this generated illustrate how the role of expert witness has drifted away from its traditional mooring. Unclarities on both sides of the bench with regard to the difference between subjective and objective probability combined with fallacies of probabilistic reasoning give pause to those who might expect a rapid entry of risk analysis into tort law and jurisprudence.

INTRODUCTION

One of the authors was recently involved as technical consultant in a civil suit described below, in which risk analysis was applied to help adjudicate damage claims. Two trains moving in opposing directions passed each other near a rural rail-road crossing. The first train passed the crossing without incident, but the second train collided with a trailer pulled by a tractor at the crossing. There were no personal injuries but the train derailed causing considerable material damage. The tractor driver, corroborated by a second witness, testified that the signalling system showed 'all clear' after passage of the first train. The railway company performed a risk analysis on the basis of which it held the tractor driver responsible for the accident and sued for damages. The driver's insurance company also acquired technical advice and submitted legal briefs. Confronted with apparently conflicting technical advice, the court convened a panel of three experts to decide the technical issues surrounding the case. Members of the Department of Mathematics of the Delft University of Technology were contacted by one of the parties to respond to some issues in the report of the expert panel.

While the damages in this case are considerable, it is nevertheless of a more 'routine' character than many of the high profile risk analysis cases which have shaped the discipline. Indeed, accidents of comparable magnitude happen almost monthly. If the tools and methods applied in this case pass muster, then the vast terrain of tort law lies open to the risk analyst.

In our opinion this case raises profound questions regarding the use of risk analysis in the adjudication of damage claims. These questions concern a threefold distinction between the adversarial method in jurisprudence, the advocacy method in science, and the traditional role of an expert-witness. Members of the legal and scientific communities do not always appreciate these differences fully. The scientific and legal methods for discovering truth must interact and must eventually reach equilibrium, but the interaction is very complex. Our conclusion is that a 'risk analysis expert witness' *cannot* play the same role in a courtroom as, say, with a ballistics expert. Indeed, the latter's testimony does not invoke probabilistic reasoning to any significant degree. If the risk analyst has a meaningful role to play in a courtroom, then the notion of 'technical expert' must be broadened.

Section two describes the uncontested facts and the preliminary judicial proceedings. Section three describes efforts by the court to invoke expert testimony. Section four comments on the expert testimony, and the final section draws general conclusions regarding the use of risk analysis in legal proceedings of this sort. The documents involved are or will be in the public domain, but the parties involved are kept anonymous, as the case is still pending. The incident occurred in 1988 in Europe, in a country counting over 1500 railway crossings and having intensive railway traffic.

UNCONTESTED FACTS AND PRELIMINARY PROCEEDINGS

Railway crossings with public roads are protected by an 'Automatic Warning System'(AWS). An approaching train is normally detected through electrical contact between the wheels and the rails over a detection zone of 1200 meters, which a passenger train travelling 140 km/hr passes in ca. 30 seconds. Detection cause the AWS to blink red lights. When the train passes, white lights blink signalling that the way is free. (At busy crossings there is also a boom which is lowered when the red lights blink, and raised when the way is free.) The AWS may fail to blink red due to poor wheel-rail contact caused by leaves or debris on the rails. Each train is equipped with an ‘Automatic Breaking Security’ (ABS) system. If poor wheel-rail contact occurs in the detection zone, the ABS issues a yellow warning light in the conductor’s cabin and sounds a warning bell. If the conductor does not respond to these signals, the train automatically breaks.

In November 1988 Mr. Hill approached a rural crossing (without boom) driving his tractor with trailer. There were no other vehicles at the intersection. Seeing the red blinking light, he stopped. Mr. Hill testified that after the first train passed, the red light stopped and the white light blinked signalling that the way was free. He started to cross the intersection and his trailer was hit by the second train approaching from the other direction. A second witness in a nearby house also testified that the light was white when Mr. Hill started to cross the intersection. The conductor of the second train testified that the ABS did not issue a warning. After the accident, the AWS and ABS were examined and found to be in working order.

After the accident, the railway company commissioned a risk analysis to determine the probability that the AWS should blink white after the passage of the first train and before the passage of the second train while the ABS, though in working order, failed to warn the conductor of the second train. The probability of this sort of AWS-ABS failure was found in 1994 to be 1 in 500,000,000 (in 1998 this probability had become 1 in 1,800,000). Mr. Hill’s insurance company commissioned studies which proposed various possible causes of AWS-ABS failure. In addition, these studies found five other cases since 1982 where the AWS failed due to poor wheel-rail contact. Three of these incidents involved the same crossing where the accident occurred. One of these three incidents occurred the previous year, in which an auto was hit by a second train under circumstances similar to the case at hand.

EXPERT TESTIMONY SOLICITED BY THE COURT

In view of the conflicting reports from the contesting parties, the court decided in 1995 to convene a panel of three experts. The panel was charged to address the following questions:

a) *In view of the uncontested facts and other relevant findings, is it theoretically possible that the AWS blinked white when Mr. Hill started to cross the intersection after the passage of the first train?*
b) *If so, under what circumstances could this occur?*
c) *Is it possible to determine after the fact with certainty whether this occurred?*
d) *What is the probability that the events in (a) should occur in the case at hand (whereby all possible causes mentioned by the contesting parties should be considered)?*

In 1998 the expert panel submitted its report. One of the three experts withdrew for reasons which themselves are now in contention. A second expert declared himself unable to address question (d) (personal communication). The remaining expert was the expert initially commissioned by the rail company.

The report found that the AWS could theoretically fail due to debris in the rails. The intersection is in a wooded area and nearby there are two private crossings with unpaved roads frequently used by agricultural vehicles. Fallen leaves and dirt could have caused poor wheel rail contact. Further, if the wheel rail contact was temporarily restored, it might happen that the ABS would give no warning in the cabin while the AWS blinked white after the first train. It is impossible, after the fact, to determine with certainty whether this occurred.

With regard to the probability of occurrence, the expert panel report states:

“ Because a statistical basis for determining the probability of failing to announce a train is not available, this probability must be estimated. Based on:

1. the assumption that in [this country] once per year a crossing system [AWS] fails to announce an approaching due to poor wheel-rail contact,
2. the fact that the probability of a second train in the detection zone of a crossing at the moment that the first train passes the crossing is 1/30, it may be expected that a second train would not be announced is once in thirty years.

The dangerous period by the approach of a second train is 1/6 of the total annunciation time of 30 seconds. This dangerous situation must therefore be expected once in 180 years, for all crossings in [this country].”

The ‘dangerous period’ is the period in which a motorist would not yet expect a second train, after the passage of the first, and would not realise that a second train is approaching. This period is stated to be 5 seconds, or 1/6 of the total annunciation time (time in the detection zone) of 30 seconds. After the dangerous period, the motorist would notice the second train, even if the blinking light inappropriately changed to white.

The final conclusion of the expert panel report is:

“Proceeding from the assumption that once in 180 years, a dangerous situation occurs, because the AWS gives a white light and a second train approaches; the probability that this happens simultaneously with the showing of a green sign in the cabin, whereby this failure would be unnoticed is a factor 10 to 1000 smaller. The probability that this happens and that the conductor also is inattentive for 5 seconds, is of course much smaller.”

It is assumed that the conductor would break manually if he saw a vehicle on the crossing, and that this would *not* happen only if the conductor were inattentive during the 5 seconds when manual breaking could avert the accident. The probability

1/30 is calculated based on a uniform distribution of trains in this country. The assumption that the AWS fails once per year in this country is defended by stating "…the number of reports to the railway company of AWS failure is minimal".

COMMENTS ON THE EXPERT PANEL REPORT

Let us first note that the expert panel's report falls well within what passes for standard practice within risk analysis. The reasoning employed is no different in form or substance than that employed in countless other risk studies. Least we forget, a paradigmatic piece of reasoning from the Rasmussen Report is reproduced in this footnote[1].

There is one major remark to be made regarding the panel's response to question (d). The report does not answer the question posed by the court. The question concerns 'the case at hand' taking into account the uncontested facts and "all possible causes mentioned by the contesting parties". Although the question is clear from a juridical standpoint, it is not phrased in technical language. Hence, the experts do not appreciate that they should give the probability of an AWS-ABS failure *given* the uncontested facts, that is a *conditional* probability. One of the uncontested facts is that the second train is in the detection zone while the first passes the intersection. Since that is given, it is clearly wrong to include the probability of this event in the calculation of AWS-ABS failure given the uncontested facts. A similar remark applies to the factor 1/6 as the fraction of 30 seconds that the motorist would not notice the second train approaching – the uncontested fact is that Mr. Hill did *not* notice its approach before starting to cross.[2]

To drive the point home, we could also compute the probability that Mr. Hill was born on 2-4-50, that his license plate was 76-KJ-32, etc and thus make the probability of the event which occurred in November 1988 appear ever smaller. Incorrect conditionalization is one of the familiar pitfalls of probabilistic reasoning.

Further, the court asks for *the* probability. The report acknowledges that an objective ("statistical") basis for the assessment is not available. They could simply have stopped there. The panel's answer to question (c) is: it is not possible to determine with certainty whether the circumstances obtained. So too, they might have replied to (d) that it is not possible to determine the probability of occurrence with certainty.

The court does not ask the panel to estimate a probability, they ask 'what is *the* probability'. Nevertheless, by juxtaposing (c) and (d) the court seems to be asking 'well, if you can't tell for certain whether it happened, how likely do you think it is that this happened?' Read in this, way the court would be asking for the experts' subjective degree of belief in AWS-ABS failure in the case at hand. Of course, in this case the court should not call this quantity *the* probability, but *your subjective probability*. In our opinion, the court's question is ambiguous, and may be interpreted as asking either for an objective probability or for the experts' subjective degree of belief. Failure to distinguish objective and subjective probabilities is another well known fallacy of probabilistic reasoning.

In any event, the panel goes on to "estimate" *the* probability. From the panel report, it is clear that their estimate is subjective. In depends on assumptions (regarding one AWS failure per year and a dangerous period of 5 seconds) which are not self-evident, are not substantiated, are not imposed by the court and are not necessarily shared by the other party. The factor "10 to 1000" is not given any further elaboration. The panel's "estimate" is the subjective probability (degree of belief) of one or at most two experts. The estimate is expressed in physical units: *number of occurrences per year*. These are the units for a frequency, not for a probability; indeed, the number of occurrences per year may be greater than one, whereas a probability is always between 0 and 1. It is well known that small *expected* frequencies approximate probabilities in some cases[3] – so well known that practitioners become very sloppy in distinguishing the two. In this case, the habitual sloppiness leads to a more serious problem: the panel fails to specify explicitly the reference class. Within which class we are counting the yearly number of occurrences? This class might be:

- all road-rail crossings in the countries with similar transport characteristics
- all road-rail crossing is this country
- all rural road-rail crossings in this country (without boom)
- all rural road-rail crossings in this country in wooded area near private unpaved crossings.
- all rural road-rail crossings in this country in wooded area near private unpaved crossings with leaves and dirt on the rails
- etc

[1] "The analysis of potential core melt accidents indicates that generally two systems would have to fail to produce such an event. If the failure probability of any system is 0.001 then the probability that two would fail could be as low as 10^{-6}. However, it is reasonable to expect that, in some cases, the use of independent failure probabilities that would be implicit in a value of 10^{-6} may not be true. The potential for common mode failures between these systems should tend in general to increase the likelihood of their failure. On the other hand, if the systems were to have totally dependent failure probabilities, then their value would be 10^{-3}. Since neither of these extremes is likely, in the absence of more precise information a reasonable value for their joint failure would be the log normal mean, or $3 \times 10^{-5} \pm$ a factor of 30."(WASH 1400, p. 68).

[2] The model which led to the probability of once in 180 years was strongly criticized in the mathematical review of the expert panel report. The essential assumption in the model was that the points at which trains pass are uniformly distributed between stations.

[3] Specifically, for a Poisson process with parameter time and rate λ, if λ is small than the probability of one or more event in one time unit is approximately equal to the expected number of events in one time unit.

It is not at all clear that the yearly relative frequency of AWS-ABS failure should be the same in each of these classes. The reference class must be chosen with care and the choice must be open to critical scrutiny. The intensity of rail and road traffic may also impinge on the choice of reference class, but may also be taken into account by changing the parameter 'time' to some appropriate unit of traffic intensity.

The subjectivist or Bayesian point of view has been embraced so whole heartedly in the risk analysis community, that many practioners may have forgotten, indeed may not know, that there are other points of view as well. An appendix is included to refresh this issue.

The objective probability[4] of a given event may be characterised as the (limit) relative frequency of occurrence of events within a reference class, from which the event in question may be regarded as a *random* sample. The unit of relative frequency (with parameter time) is: *Number of occurrences, per unit time divided by number in the reference class* (assuming the number in the reference class is constant in time). Hence, using frequency units [#occurrences/yr.] instead of relative frequency units [(#occurrences / #reference class)/yr.] allows the reference class to remain unspecified and thus unmotivated.

Of course, the panel specifies a reference class implicitly by considering the number of occurrences in this country, per year. However, taking this as the reference class for the "the case at hand", accident of Mr. Hill, is problematic. In the case at hand, there is the likelihood of fallen leaves and dirt on the rails, there is a history of similar incidents, and there is testimony of a witness corroborating the testimony of Mr. Hill that the light was white when he started over the crossing. Each of these facts might be deemed relevant for defining the reference class. In other words, each of these facts might make Mr. Hill's crossing a *non*random sample. Here again, we may conclude that the panel fails to answer the question posed by the court, in so far as they (implicitly) select a reference class which is not self-evidently appropriate for "the case at hand".

Confusing frequencies and relative frequencies, and improper or unclear specification of the reference class are again familiar pitfalls in probabilistic reasoning.

RISK ANALYSIS AND JURISPRUDENCE

The adversarial method of adjudication may be characterised as follows: opposing parties try to convince an impartial party (the judges or a jury) of their view, according to rules laid down by the court to insure a 'level playing field'. The interested parties are not expected to be 'fair' or 'unbiased' and arguments ad hominum are freely admitted. Fairness, and even truth, are supposed to result from the rules under which the contest transpires.

The scientific method is advocatory, as opposed to adversarial. Advocates of different viewpoints try to convince an impartial body (the scientific community) of their views, according to rules laid down by the scientific method. Membership in the community is defined inter alia by shared values, including honesty, open-mindedness and the like; and advocates, as members of the community, are expected to uphold these values. Arguments ad hominum are taboo. Nonetheless, advocacy does entail a certain interestedness as reflected e.g. in what sort of evidence to gather, what sort of experiments to perform, what sort of theories to develop and test.

In its efforts to ensure a level playing field, the court, as in the case of Mr. Hill, may solicit advice from expert witnesses. The (idealised) traditional expert witness confines himself to statements on which the entire scientific community would agree; he does not expound personal conjectures, untested hypotheses, disputed theories and the like. His statements are therefore non-probabilistic. Put simply, the court may direct yes-no questions to the expert, and the expert answers either 'yes' or 'no', or declines to answer. Moreover, any other expert would give the same answers, since they are merely reporting the consensus of the community.

When an expert witness is asked to state *objective* probabilities, he is still, in principle, within his traditional role. The question should make clear *what* probability is queried, i.e. what is the reference class, and the expert should answer by giving the ratio of the number of occurrences to the number in the class, perhaps with a confidence interval reflecting sampling fluctuations. Every expert would give the same answer.

When an expert witness is asked to give *subjective* probabilities, or to 'estimate' objective probabilities with (his) degree of belief, then we are clearly outside the traditional role of an expert witness. It is no longer the case that all experts would give the same answer. Experts are no longer informing the court of the community consensus; rather they, in principle, are appearing as advocates.

This does not mean that expert-advocates have no place in the courtroom. However, it is essential that all parties appreciate the difference between the traditional expert witness and the expert-advocate. The single most significant difference is this: unlike the traditional expert witness confining his testimony to views shared by the scientific community, expert-advocates will give personal views and will not all say the same thing.

Risk analysis finds itself at the boundary of these two roles because it deals with probabilistic reasoning and has explicitly endorsed the use of objective *and* subjective probabilities in the

[4] The literature on this subject is vast, the characterization given here is necessary but not sufficient. Such characterizations require an antecedent definition of 'random sample'. The appendix introduces the interested reader to the literature.

form of expert judgement. Expert judgement, by its nature, is not consensual. Different experts will have different judgements – otherwise we would not be speaking of expert *judgement*.

If the example discussed here is representative, then it is fair to say that confusion can easily arise, on both sides of the bench, with regard to whether an expert witness is within his traditional role (reporting the community consensus) or within an advocacy role (reporting his particular point of view). The previous section identified some common pitfalls in probabilistic reasoning, namely:

- ♦ Incorrect conditionalization
- ♦ Confusing objective and subjective probabilities
- ♦ Confusing frequencies and relative frequencies
- ♦ Improper/unclear identification of the reference class.

The example of Mr. Hill's accident illustrates how these pitfalls can create opportunities for confusion. Thus we saw that the expert panel did not answer the question (d) posed by the bench at all, but replaced this with a different question (by incorrect conditionalization). They also fail to fully address "the case at hand" by (implicitly) choosing a problematic reference class by dint of sloppiness with regard to frequencies and relative frequencies. The bench contributed to all this by not stating clearly whether they were interested in an objective or a subjective probability. The fact that mathematical consultants are called in at all, indicates that substantial confusion exists on both sides of the bench. There is no evidence, and no reason to believe, that all this was done deliberately. Probabilistic reasoning is sufficiently subtle and unfamiliar for experts and non-experts alike to fully explain the difficulties encountered in this case.

This is not the place to discuss methods for factoring expert judgement into jurisprudence. Suffice to say that the question how to use expert judgement in decision science and risk analysis is still under active discussion (Cooke, 19912, Harper et al 1995, USNRC 1997).

APPENDIX: FOUNDATIONS REFRESHER

A 'representation of uncertainty' is specified by giving (a) a set of axioms, (b) interpretations of the primitive terms in these axioms (i.e. terms which are not defined in terms of other formal terms), and (c) measurement procedures implementing the interpretations in practice (Bedford and Cooke 2000, chapter 1, Cooke 1991, appendix A, van Lambalgen 1987). Within the risk community, most people would agree that Kolmogorov's axioms describe the formal properties of uncertainty, namely as a positive normed measure.

Of the many interpretations which have been proposed, two have survived their original proponents, namely the subjectivist and the objectivist or frequentist interpretations. According to the subjectivists probability is a primitive term. It is interpreted as degree of belief of a rational individual. A rational individual is one whose preferences satisfy certain axioms. One then proves that for a rational agent, rational preference can be represented as expected utility: There exists a unique probability over the set of possible worlds, and an affine unique (i.e. unique up to a choice of 0 and unit) utility function on the set of consequences, such that for any options A and B, A is preferred above B if and only the expected utility of A is greater than that of B.

According to the frequency interpretation, probability is *not* a primitive formal term, rather probability is defined as limiting relative frequency in a *random* sequence. Randomness is the primitive term which must be interpreted. Although this interpretation was launched at the beginning of the 20th century, a satisfactory operational definition of randomness was not found until the 1960's (see van Lambalgen 1987). To give the drift of the modern discussions, we focus on sequences of 0's and 1's in which the '1' should 'have probability ½'. For one thing, this means that the limiting relative frequency of '1' should be ½. Consider two sequences

0,1,0,1,0,1,0,1,…
1,0,0,1,1,1,0,1,…

The first is regularly alternating, the second is the parity of ciphers in the decimal expansion of e. In both the limiting relative frequency of '1' is ½. The second would be considered random but the first would not. Although it has the desired limit relative frequency, it does not possess other properties which we require of random sequences. One way of formalizing this is to give a test statistic with which we would reject the hypothesis that the outcomes were generated by independent tosses of a fair coin at any finite level of significance. One such statistic might be: the relative frequency of '1''s following a '1'. The probability of seeing no '1's following a '1' in any finite initial sequence goes quickly to zero on the above hypothesis. Interpreting randomness comes down to specifying a set of statistical tests which a putatively random sequence must satisfy. While it may seem hopeless to draw up a definitive list of all such tests, logicians have ways of capturing these lists without actually drawing them up. Hence, one says that random sequences are those whose finite initial subsequences pass all recursive statistical tests. That includes all tests you will ever think up.

Summarizing, the statement *the probability of heads with this coin is ½* is interpreted as follows:

The finite initial sequences of outcomes of tosses with this coin pass all recursive tests for randomness, with limiting relative frequency of '1' = ½.

An infinite sequence of 0's and 1's may be regarded as a real number in binary expansion. One can prove that 'almost all real numbers in the unit inverval are random'. More precisely the set of sequences which do not pass all recursive tests for randomness with limiting frequency r is a set of measure zero with respect to the product measure generated by r.

Although the terms subjectivist and Bayesian are often used interchangeably, it is more accurate to distinguish them. Subjectivism is a position with regard to the interpretation of the probability formalism, Bayesianism is a theory of statistical inference. This type of inference is most convenient within the subjectivist interpretation, but can of course be used within the frequency interpretation as well. In this case one sometimes speaks of 'Empirical Bayesians'.

ACKNOWLEDGMENTS

The authors gratefully acknowledge useful discussions with Piet Groeneboom, Corine Reedijk and Linda Merrienboer.

REFERENCES

Cooke, R. M. (1991) *Experts in Uncertainty*, Oxford University Press.

US Nuclear Regulatory Commission (1975) *Reactor Safety Study* WASH-1400, U.S.N.R.C. NUREG-751014.

F.T. Harper, L.H.J. Goossens, R.M. Cooke, S.C. Hora, M.L. Young, J. Päsler-Sauer, L.A. Miller, B. Kraan, C. Lui, M.D. McKay, J.C. Helton and J.A. Jones, (1995) *Joint USNRC/CEC consequence uncertainty study: Summary of objectives, approach, application, and results for the dispersion and deposition uncertainty assessment,* Prepared for U.S. Nuclear Regulatory Commission and Commission of European Communities, NUREG/CR-6244, EUR 15855 EN, SAND94-1453, Washington/USA, and Brussels-Luxembourg, (Vol. 1-3).

Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts, NUREG/CR-6372, (1997) U.S. Nuclear Regulatory Commission.

Van Lambalgen, M. (1987) *Random Sequences* PhD dissertation, University of Amsterdam.

Bedford, T. and Cooke, R. (2000) *Mathematical Tools for Probabilistic Risk Analysis,* Cambridge University Press.