Legal knowledge based systems
JURIX 94
The Relation with Legal Theory

The Foundation for Legal Knowledge Systems
Editors:
H. Prakken
A.J. Muntjewerff
A. Soeteman

D. Hamer and D. Hunter, *Probability and Story-Telling: Normative and Descriptive Models of Juridical Proof,* in: A. Soeteman (eds.), Legal knowledge based systems JURIX 94: The Foundation for Legal Knowledge Systems, Lelystad: Koninklijke Vermande, 1994, pp. 93-104, ISBN 90 5458 190 5.

More information about the JURIX foundation and its activities can be obtained by contacting the JURIX secretariat:

Mr. C.N.J. de Vey Mestdagh
University of Groningen, Faculty of Law
Oude Kijk in 't Jatstraat 26
P.O. Box 716
9700 AS Groningen
Tel: +31 50 3635790/5433
Fax: +31 50 3635603
Email: sesam@rechten.rug.nl

Probability and Story–Telling: Normative and Descriptive Models of Juridical Proof

David Hamer
Director of Studies
Australian College of Law
Melbourne, Australia

Dan Hunter
CLAIR Project
Law School
University of Melbourne
Parkville Victoria Australia

Abstract

The operation of the law rests on the selection of an account of the facts. Whether this involves prediction or postdiction, it is not possible to achieve certainty. Any attempt to model the operation of the law completely will therefore raise questions of how to model the process of proof. In the selection of a model a crucial question will be whether the model is to be used normatively or descriptively. Focussing on postdiction, this paper presents and contrasts the mathematical model with the story model. The former carries the normative stamp of scientific approval, whereas the latter has been developed by experimental psychologists to describe how humans reason. Neil Cohen's attempt to use a mathematical model descriptively provides an illustration of the dangers in not clearly setting this parameter of the modelling process. It should be kept in mind that the labels 'normative' and 'descriptive' are not eternal. The mathematical model has its normative limits, beyond which we may need to critically assess models with descriptive origins.

1 Introduction

Law is an instrument for ordering society. It intervenes in the resolution of disputes about past events, and prescribes rights, duties and obligations into the future. In both cases, law is applied to what are commonly described as 'facts', but they are not facts in the true sense of the word. They are merely versions of facts about which the actor has a degree of belief which is less than certainty. The law's application therefore rests on probability judgments.

This paper focuses on the way in which judges and juries view the proof of past events, postdiction, although much of the discussion will be of relevance to prediction also. It may be thought that certainty is possible about some events in the past. For example, most people would be certain about their date of birth. But would this certainty be justified? Certainly, they were there at the time, but it is unlikely that they have a clear recollection of it. Certainly, one of their parents was there, but how clear is their recollection? There is a document with the date of birth on it, but can we be certain that this is the correct date? Perhaps the date was typed incorrectly... If there could be doubts about this most keenly witnessed and documented of events, *a fortiori* proof of more *ad hoc* events will be problematic. In resolving a dispute about a car accident, or assessing the evidence in a murder prosecution, one of the court's most difficult tasks will be 'finding' the 'facts'.

If the law's operation rests on such an unsteady foundation, then attempts to model the law in isolation will be incomplete. Research in artificial intelligence and law has not yet examined probabilistic reasoning, though probabilistic reasoning is a source of significant work in other domains [Pearl, 1988]. In time the artificial intelligence and law movement will be need to deal with the proof issue. Perhaps, with the numerical basis of computers, and their inordinate capacity for calculation, the natural tendency would be to adopt a mathematical probabilistic model. However first the question should be asked whether the system is to be descriptive or normative. In either case, a mathematical model may not be the most appropriate.

The next section of this paper examines the findings of experimental psychologists which reveal that human judgments of uncertainty do not follow the rules of mathematical probability. The story model, on the other hand, provides a clear description of the human process. In the third section support is found for the story model in alternative formulations of the civil standard of proof., where there exists a tension between alternative formulations. The implications of the mismatch between mathematics and human probability judgments for the modelling process are discussed. In the fourth section, a variation on the mathematical model of proof is discussed – Neil Cohen's confidence model. This paper argues that the statistical concepts on which the model rests are meaningless in non-statistical cases. The model however illustrates the risks of not being precise about the goals of the modelling process. The final section explores the relationship between the concepts 'descriptive' and 'normative'. The normative limits of the mathematical model are emphasised and suggestions are made for future research.

2 Human probability judgments

This section examines work by experimental psychologists investigating human probability judgment. These and other experiments reveal that people do not assess evidence mathematically, but employ a number of simplifying processes, or heuristics. "In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors." [Tversky & Kahneman, 1982a].

2.1 Experiment 1: Direct identification evidence and incidental base rate data

In [Tversky & Kahneman, 1982b, p. 3] the subjects were given the following information: A cab was involved in a hit and run accident at night. Eighty-five per cent of the cabs in the city that night were green, and 15 per cent were blue. A witness identified the cab as blue. The reliability of the witness was tested in conditions similar to those of the sighting and it was discovered that they gave a correct identification 80 per cent of the time. The cab ownership figures can be described as incidental base rate data. They are termed 'base rate' because they have equal application to all the cab accidents in the town. They are incidental in that there is no causal link with the accident; they provide no explanation for why the accident occurred. The witnesses identification is direct in that, if accepted, it provides a conclusive answer to the question in point.

The subjects were asked to give the probability that the cab involved in the accident was blue rather than green. Most subjects said that there was a probability of 80 per cent that the cab was blue. They apparently took account only of the reliability of the witness, and ignored the cab ownership data.

Mathematical probability theory indicates that there is a 41 per cent probability of the cab being blue. This may be calculated as follows. Imagine that the witness observed one hundred such accidents. One would expect that 85 of the accidents were caused by green cabs, and that 15 accidents were caused by blue cabs. Out of the 85 green accidents, the witness would be expected to make 68 correct identifications of the cab as green (80 per cent of 85), and 17 incorrect identifications of the cab as blue (20 per cent of 85). Out of the 15 blue accidents, the witness would be expected to make 12 correct identifications of the cab as blue (80 per cent of 15), and 3 incorrect identifications of the cab as green (20 per cent of 15). Hence one would expect the witness to identify a cab as blue in 29 cases (12 plus 17), out of which one would expect 12 identifications to be correct. Therefore the probability of a blue identification being correct is 41 per cent (12 divided by 29). This is an application of Bayes theorem.

A common objection to this solution of the problem is that the cab ownership figures have nothing to do with the accuracy of the witness [Cohen, 1981, p. 328]. This is quite right. But the cab ownership figures are relevant to the credibility of the witness when they make the unlikely report that the cab was blue [Mackie, 1981, p. 346]. Another common objection is that the cab ownership figures are too remote to be relevant. The figures we need are of the accident rates of the two cab companies [Cohen, 1981, pp. 329,365; Brook 1985, p. 346; Callen 1982, p. 8n33]. There is some truth in this too. Accident rate figures would provide a better base rate. But if they are unavailable, the cab ownership figures should be used. To disregard the cab ownership figures, is effectively to assume that both companies have an equal number of accidents. And this is to assume that each blue cab is 5.67 times as likely to have an accident as each green cab. This seems a dangerous assumption if there is no evidence to support it.

It is important to note that this process of interpreting and applying the statistics is not, in itself, a mathematical process, and, even in this carefully designed experiment, the issues raised are far from trivial. Jonathan Cohen and Callen, while strongly opposed to mathematical theory, would not argue with the validity of the calculation, but rather would question its relevance to the real world. Mathematics offers the promise of 'correct' answers, but within what limits? We will return to this issue in the conclusion of this paper.

2.2 Experiment 2: Direct identification evidence and causal base rate data

In this experiment [Tversky & Kahneman, 1982b, p. 157] the subjects were told that while the two cab companies were roughly equal in size, 85 per cent of the accidents involved green cabs, and 15 per cent involved blue cabs. Again the subjects were told that a witness with 80 per cent reliability identified the cab as being blue. The only distinction between this experiment and the first is that the incidental base rate data was replaced by causal base rate data. The accident statistics are still base rate data in that they have equal application to each accident. However they have a causal connection with the accident. Factors lying behind past accidents appearing in the statistics arguably lie also behind the present accident. The mathematical solution to this problem is the same as in the first, however one might expect less argument about the relevance of the answer, given the causal connection between the statistics and the present accident.

In this experiment the answers were highly variable, but the average response was that there was a 60 per cent probability that the cab was blue. Interestingly, this lies about half way between the reliability of the witness (80 per cent) and the correct answer (41 per cent). This experiment when considered in conjunction with the first reinforces the hypothesis that humans have a strong preference for direct evidence over base rate data. We may further hypothesise that they have a preference for causal base rate data over incidental base rate data.

2.3 Experiment 3: Circumstantial identification evidence and incidental base rate data

In this experiment [Tversky & Kahneman, 1982b, p. 158], as in the first, the subjects were told that 85 per cent of the town's cabs were green, and 15 per cent were blue. However, the direct identification evidence of the eyewitness was replaced by circumstantial identification evidence. The subjects were told that mobile–phones had been installed in 80 per cent of green cabs and 20 per cent of blue cabs, and that the cab involved in the accident was equipped with a mobile-phone. This evidence is circumstantial in that, even if accepted, it does not conclusively establish the colour of the cab in the accident.

In this experiment, the subjects utilised the incidental base rate data to a considerable degree. The median answer was that there was a 48 per cent probability that the cab in the accident was blue, which is quite close to the answer generated mathematically (41 per cent).

A comparison between this experiment and the first suggests the hypothesis that humans have a preference for direct identification evidence over circumstantial identification evidence.

2.4 Experiment 4: Incidental base rate data

In a fourth experiment subjects were only given the cab ownership figures [Tversky & Kahneman, 1982b, p. 157]. Kahneman and Tversky found that in the absence of other kinds of evidence, almost all subjects utilised the incidental base rate data and arrived at the correct answer.

2.5 Interpretation of the results: the story model

Psychologists suggest that 'when estimating probabilities people can use several cognitive strategies based upon different types of information.' [Hendrickx *et al.*, 1989, p. 41]. Two that have been identified are frequency-based judgments, and scenario-based judgments. Frequency-based judgments are, as the name suggests, ideally suited for repetitive activities [Hendrickx *et al.*, 1989, p. 60]. Scenario-based judgments, on the other hand, are made where the events are perceived as so unique that past history does not seem relevant to the evaluation of their likelihood [Tversky & Kahneman, 1982c, p. 177]. In most everyday situations where people make probability judgments, frequency data will be unavailable. [Hendrickx *et al.*, 1989, p. 58] indicate that people have a strong leaning towards scenario-based judgments, utilising frequency data only when scenario data is unavailable. [Kahneman & Tversky, 1982c, p. 176] suggest that typically people "evaluate likelihood by attempting to construct stories, or scenarios. ... The plausibility of such stories, or the ease with which they come to mind, can provide a basis for the judgment of likelihood." These commentators warn that this method can lead to systematic biases. [Slovic *et al.*, 1976, p. 178] suggest that "scenarios which tell a 'good story' by burying weak links in masses of coherent detail may be accorded much more credibility than they deserve." Indeed the subject may ignore the credibility of the source of the evidence, and focus instead on the inherent credibility of the story which the evidence tells.[1] A corollary of the story model is that mathematically strong data may be considerably discounted if it is offers little colour or detail to a scenario.

This account of human probability assessment offers an explanation for the experimental results described above. The best story of the cab accident will be told by the eyewitness. They were here at the time, are an actor in the story of the accident, and may be able to tell not only who was involved, but also when, where, how and why the accident occurred. Because of the detail that the witness is able to offer, the fact–finder gives their evidence considerable weight. By comparison, the cab ownership statistics in the first experiment add very little detail to the account of the accident. At best, they provide a clue as to who may have been involved, and say nothing about where, when or how the accident occurred.

The accident statistics of the two cab companies in the second experiment have more to say than the raw cab ownership data. In addition to telling who may have been involved, they tell why the accident occurred. They point to the poor driving ability of the green cab drivers as the possible cause of the accident. However they still offer little detail in comparison with the evidence of the eyewitness.

The circumstantial identification evidence in the third experiment also adds some colour to the story of the accident. In addition to suggesting which company's cab was involved, this evidence may tell when or where the accident occurred. For example, if the evidence were in the form of remnants of a mobile-phone at the scene of the accident, it would at least tell where the accident occurred. If the evidence were in the form of a message transmitted from the accident–cab, it may indicate when the accident occurred. Unlike the base rate data, the evidence is unique to this accident and so may be given

[1] One of the referees of this paper made this point with reference to the work of the Dutch scholars, Wagenaar, Cohen and Koppen. We are grateful for this reference, however were unable to find it, or any similar work by this group, in English translation.

greater weight by fact–finders.

While the fourth experiment suggests that people will employ a mathematical frequency method of judgment where scenario information is unavailable, the other experiments show a human preference for a scenario-based process of judgment.

3 Human probability judgments and the civil standard of proof

Most law trials will arise out of a unique series of events. One may view certain mass–tort cases as non–unique, the gulf war syndrome is perhaps the most recent example, and it is not a coincidence that statistical evidence is often the most useful evidence in these cases. But generally speaking, statistical evidence will be unavailable. When it is available, in view of the preference humans have for scenario-based judgments over frequency-based judgments, statistical data is likely to be considered less persuasive than scenario data [Saks & Kidd, 1980, p. 148], but see [Tribe, 1971, p. 1334,1376].

It is not surprising then that the parties do not appeal to the fact–finder's frequency-based judgment faculty. The parties instead seek to tell the fact–finder their story of what happened. How do they make their story convincing? [Binder & Bergman, 1994] give this practical advice:

> "Details...are the veritable lifeblood of persuasive stories...(T)he details make the story vivid; they allow the fact–finder to visualize the actual events. Hence details provide specific evidence and make stories credible and persuasive."

The psychologists [Pennington & Hastie, 1991, p. 520] in their attempt to "develop a scientific description of the mind of the juror as it is revealed in the decision making process" found that "a central cognitive process...is *story construction*". Evidence theorists are arriving at the same conclusion [Allen, 1991a; La Rue, 1992].

The story model provides insights into the now notorious risks inherent in eyewitness identifications: persuasive to human factfinders but frequently unreliable [Loftus, 1979]. It also provides a new perspective on a series of cases decided by the High Court of Australia arising out of collisions. At trial there were no credible eyewitnesses and the evidence consisted of little more than "the position and state of the vehicles after the collision.".[2] Could this satisfy the civil standard of proof?

Traditionally it is said that the plaintiff in a civil case must establish their facts 'on the balance of probabilities' or by 'a preponderance of evidence' [Byrne & Heydon, 1986; Eggleston, 1983]. These expressions lend themselves to a mathematical interpretation, that the plaintiff must establish their case to a mathematical probability of greater than 50 per cent. Often judges have described the civil standard as requiring the plaintiff to establish that their version of the facts is 'more probable than not.'[3]

[2]TNT v. Brooks (1979) 23 A.L.R. 345, 350 (per Gibbs J.); Jones v. Dunkel (1959) 101 C.L.R. 298, 306 (per Kitto J.). See also West v. Government Insurance Office of New South Wales (1981) 148 C.L.R. 62; Goodwin v. Nominal Defendant (1979) 54 A.L.J.R. 84; Holloway v. McFeeters (1956) 94 C.L.R. 470; Luxton v. Vines (1952) 85 C.L.R. 352; Bradshaw v. McEwans Pty Ltd (1951) unreported.

[3]E.g. TNT v. Brooks (1979) 23 ALR 345 at 351-54 (per Murphy J); Goodwin v. Nominal Defendant (1979) 54 A.L.J.R. 84 at 86 (per Stephen, Mason, Aickin and Wilson JJ); Bradshaw v. McEwans Pty Ltd (1951) unreported (full court) applied in Holloway v. McFeeters (1956) 94 C.L.R. 470 at 480-81 (per Williams, Webb and Taylor JJ.). Cf. Davies v. Taylor [1974] A.C. 207 at 219 (per Lord Simon).

However, in the collision cases, where the evidence was scarce and circumstantial, certain judges expressed dissatisfaction with a mathematical interpretation:

> "The truth is, that when the law requires the proof of any fact, the tribunal must feel an actual persuasion of its occurrence or existence before it can be found. It cannot be found as a result of a mere mechanical comparison of probabilities independently of any belief in its reality."[4]

These judges appear to be demanding a scenario-based proof 'special to the particular case under consideration'. A frequency-based proof, resting upon 'general considerations as to the likelihood of negligent conduct occurring in the conditions which existed at the time and place of the collision'[5] would not be sufficient. To accept the plaintiff's version without a detailed story would be 'a mere mechanical comparison of probabilities (independent) of any belief in its reality', a guess,[6] and 'purely conjectural.'[7]

It is difficult to say that, by abandoning the mathematical standard, the judges reached the wrong result. If there was a 'correct result' readily at hand the issue of proof would not have been raised. In relation to the cab hypotheticals discussed above, despite their deliberate design, the issue of how correctly to interpret the evidence was not wholly excluded. (Although it appears to us that the mathematical answer was the most correct.) When we come to the collision cases, the question of interpretation becomes larger, and mathematics does not dictate the correct answer. However, a narrower question can be asked: Was it wrong for the judges to abandon the traditional 'more probable than not' standard because of the scarcity of evidence? Despite the efforts of a number of commentators to justify a non–mathematical standard [Callen, 1982, Dant, 1988] we consider that the judges were wrong. The discussion above demonstrates that, in connection with the collision cases, the mathematical model has some normative validity, and the story model descriptive validity. These models suggest that the judges were unjustifiably sceptical of what they recognised as mathematically strong evidence.

4 Measuring the quantity of the evidence

A distinctive feature of both the cab hypotheticals and the collision cases decided by the High Court is their dearth of evidence. This feature is shared by another hypothetical, much discussed by evidence scholars, the Gatecrasher case of Jonathan Cohen [Cohen, 1977]:

> "Consider a case in which it is common ground that 499 people paid for admission to a rodeo, and that 1000 are counted on the seats, of whom A is one. Suppose no tickets were issued and there can be no testimony as to

[4] Briginshaw v. Briginshaw (1938) 60 C.L.R. 336, 361(per Dixon J.). Cf. Holloway v. McFeeters (1956) 94 C.L.R. 470, 488 (per Kitto J.); Jones v. Dunkel (1959) 101 C.L.R. 298, 304-5 (per Dixon C.J.) applied in West v. Government Insurance Office of New South Wales (1981) 148 C.L.R. 62, 66 (per Stephen, Mason, Aickin, and Wilson JJ.). Cf. Sargent v. Massachusetts Accident Company 28 N.E.2d 825 (1940), 827 (per Lummus J.) applied in Smith v. Rapid Transport 58 N.E.2d 754 (1945), 755.

[5] Jones v. Dunkel (1959) 101 C.L.R. 298, 305 (per Kitto J.).

[6] Jones v. Dunkel (1959) 101 C.L.R. 298, 305 (per Dixon CJ).

[7] TNT v. Brooks 23 A.L.R. 345, 350 (per Gibbs J; Stephen, Mason and Aicken JJ agreeing).

> whether A paid for admission or climbed over the fence. So there is a .501 probability, on the admitted facts, that he did not pay" [Kaye, 1979, p. 101].

A number of evidence commentators have suggested that further investigation of the quantity of evidence may lead to a resolution of the uncertainty surrounding the civil standard of proof [Cohen, 1977; Lempert, 1988, pp. 76–80; Birmingham & Dunham, 1987, pp. 813–814; Brilmayer, 1988, pp. 156–159]. They have, however, stumbled at the outset on the theoretical difficulty of finding a sensible measure of 'quantity of evidence' other than the traditional mathematical probability.

Neil Cohen [Cohen, 1985; Cohen, 1987] purports to have overcome this difficulty. According to Cohen, the fact–finder, when presented with a body of evidence, first determines how probable the evidence renders the plaintiff's version of facts, and secondly, by considering the quantity of evidence before them, the fact–finder measures their confidence in this determination. If they have sufficient confidence that the *true probability* exceeds the standard of proof, then the fact–finder is justified in settling on that version of facts. Cohen's theory rests on an analogy with statistical sampling methods. Suppose we have a large vat of black and white marbles. If we drew fifty marbles of which thirty were white, we would estimate that there would be a sixty per cent probability that the next marble to be drawn would be white. If we drew 100,000 marbles and 60,000 were white, we would form the same conclusion, but with much more confidence in this estimate. Statistical methods can measure the confidence in cases such as these with a great deal of accuracy.

Cohen suggests that the concept of confidence explains why the plaintiff would fail in the Gatecrasher case; while the evidence renders the plaintiff's case more probable than not, 'we cannot say with sufficient confidence that the true probability of liability exceeds 0.5.' In the Gatecrasher case, the probability of 50.1 per cent from the attendance figures is not the 'true probability', because:

> "Absent from this determination is a great deal of information beyond the overall proportion of gatecrashers in the rodeo crowd. For example, does the defendant have a witness who can testify that he or she saw the defendant purchase a ticket? What is the defendant's reputation for honesty?" [Cohen, 1985, pp. 407–408].

Cohen's theory applies quite readily to his carefully designed hypothetical cases which raise sampling problems and have purely statistical issues in dispute. His 'confidence' concept has a well recognised definition in such a context, and the 'true probability' is itself a statistic. However, Cohen gives little guidance as to how his theory applies to non-statistical cases. In a case such as the Gatecrasher case, what is the 'true probability' that Cohen alludes to? Cohen fails to adopt a consistent definition for this concept that lies at the core of his theory.

At one point Cohen suggests that the "true probability' is 'based on *all possible evidence*" [Cohen, 1985, p. 399 emphasis added]. Is Cohen referring here to *ideal evidence*, such as the testimony of a hypothetical, wholly reliable and creditworthy eyewitness? But ideal evidence in the hypotheticals, unlike Cohen's sampling hypotheticals, would provide us with certainty, rather than a probability. We would know the colour of the next ball to be drawn from the vat, not merely the probability that it would be white. As [Cohen, 1985, pp. 397n74] himself recognised at a different point, "If one knew all the possible information concerning an event, it would be meaningless to speak of probabilities. (If) one had perfect information ... one could predict [or postdict] with certainty." On this interpretation of 'true probability' Cohen's interpretation of the civil standard would merge with the traditional one-dimensional probabilistic standard. The fact–finder would simply be required to have a sufficient level of confidence that the true facts were as the evidence indicated. In the Gatecrasher case, the fact–finder would be 50.1 per cent confident that the defendant was in truth a gatecrasher. If Cohen is dissatisfied with this result he is merely advocating a higher mathematical standard of proof.

At a second point Cohen states that the 'true probability' is based upon "the totality of *all available evidence*" [Cohen, 1987, p. 85 emphasis added]. But if this is what Cohen means, his theory runs into other difficulties. If the fact–finder is not presented with this evidence how could they know its quantity and substance? Whether the 'true probability' would differ markedly from the estimate would depend upon what other available evidence the fact–finder chose to imagine. The rule would have no application to the collision cases where there was no further available evidence. One might also question whether the plaintiff should be or is in practice penalised for the absence of available evidence in the defendant's domain.[8]

A third interpretation is that Cohen is suggesting that the fact–finder measures their confidence merely by counting the number of pieces of evidence before it. Perhaps this is what Cohen meant in saying "the precision of a subjective probability derived in the legal context is a function of the *quantity of information* upon which it is based." [Cohen, 1985, p. 398, emphasis added]. This interpretation has, at least, the merit of simplicity. However it is not consistent with court practice. Fact-finders often place a great deal of confidence, sometimes too much, in a single eyewitness identification.

Cohen claims to have constructed "a more realistic probabilistic formulation of the civil standard of proof." [Cohen, 1985, p. 386]. But without a clear meaning to 'true probability' his model does nothing at all. His project was flawed from the outset by his failure to decide whether he was creating a normative or a descriptive model of the process of proof. The use of statistical methods suggests a normative application. And yet, as we have seen the statistical concepts do not readily transfer to non-statistical cases. Cohen states that judicial statements which talk of the need for an " 'actual belief' ... can be understood as unfocused attempts to describe the concept of confidence" [Cohen, 1985, p. 420] implying a descriptive model. But surely we should consider the findings of experimental psychologists if we are to build a descriptive model, and as illustrated above, these indicate that human reasoning is far from statistical.

5 Conclusion

Researchers in the area of law and artificial intelligence have not yet examined the fact finding process, and yet the uncertain adoption of a version of facts is a necessary step in the application of the law. Inevitably the issue will be faced as researchers approach the goal of complete legal systems. The first step in designing a proof system will be

[8] Jones v. Dunkel (1958-59) 101 CLR 298, 312 (per Menzies J), 321–22 (per Windeyer J); Cf. Wigmore on Evidence (Chadbourn Revision) Vol 2, 192-214.

the selection of an appropriate model.

In this paper we have presented the mathematical and story models of fact-finding, and discussed their normative and descriptive validity to a number of hypothetical and actual cases. Two conclusions from our discussion present themselves immediately. The story model has considerable descriptive power, and the mathematical model, within carefully drawn limits, offers useful normative guidance. We will conclude by commenting on the normative limits of the mathematical model, and the possible normative application of the story model.

One clear limit to mathematical theory has been pointed out. Mathematical probability theory provides rules by which a set of numerical inputs may be combined to produce a set of numerical outputs. The theory says little about the meaning of the inputs and outputs in the real world. These questions lie on the border of the mathematical realm and beyond.

A second limit to mathematical theory is implicit in our discussion. A noticeable feature of the cab hypotheticals, the gatecrasher case and the collision cases is their scarcity of evidence. It is because of this feature that mathematical theory offers a relatively clear prescription – necessary calculations are confined. Recent research into the application of mathematical theory suggests that it has questionable application to more complex cases where the number of necessary calculations explode exponentially to the point of being totally unwieldy. As Callen has noted 'for the consistent use of Bayesian theory for the updating of probabilities by conditionalisation, where thirty items of evidence are introduced relevant to an inference, one must record a billion probabilities.' [Callen, 1986, pp. 725n72], see also [Cherniak, 1986, p. 93]. Among the community of evidence scholars a number of formerly committed mathematicists are reviewing their positions. Contrast for example [Koehler & Shaviro, 1990] with [Shaviro, 1991, p. 1112]. This should sound a note of warning for researchers in law and artificial intelligence.

Where then do we look for normative guidance at the interpretative stage, and in evidentially complex cases, in the absence of mathematical theory? As in other areas of research in applied cognitive science, it appears the only way forward is to critically assess the performance of experts: to discover and model the heuristics which they employ. The story model, while subject to the normative limits noted above, will provide a useful starting point in this work. We do not advocate the replication of juridical 'mistakes' in legal expert systems. We do believe however that using the story model in these expert systems may provide insights into the human fact-finding process and improve it.

References

Allen, R.J. 1991a. The Nature of Juridical Proof. *Cardozo Law Review*, 13:373.

Allen, R.J. 1991b, On the significance of batting averages and strikeout totals: A clarification of the 'Naked Statistical Evidence' debate, the meaning of 'Evidence,' and the requirement of proof beyond a reasonable doubt, *Tulane Law Review*, 65:1092.

Binder, D. and Bergman, P. 1984. *Fact Investigation From Hypothesis to Proof.*

Birmingham, R. and Dunham, N. 1987. An Evidentiary Reading of Naked Statistical Proofs. *Saint Louis University Law Journal*, 13:797.

Brilmayer, L. 1988. Second-Order Evidence and Bayesian Logic. In Tillers, P. and Green, E.

(ed.) *Probability and Inference in the Law of Evidence: The Uses and Limits of Bayesianism*, 147.

Brook, J. 1985. The Use of Statistical Evidence of Identification in Civil Litigation: Well-Worn Hypotheticals, Real Cases, and Controversy. *Saint Louis University Law Journal*, 29:293.

Byrne, D.M. and Heydon, J.D. 1986. *Cross on Evidence*. 3rd Australian edition.

Callen, C. 1982. Notes on a Grand Illusion: Some Limits on the Use of Bayesian Theory in Evidence Law. *Indiana Law Journal*, 57:1.

Callen, C 1986. Second-Order Considerations, Weight, Sufficiency and Schema Theory: A Comment on Professor Brilmayer's Theory, *Buffalo Law Review*, 66:715.

Callen, C., 1991, Adjudication and the Appearance of Statistical Evidence, *Tulane Law Review*, 65:457.

Cherniak, C., 1986, *Minimal Rationality.*

Cohen, L.J. 1977. *The Probable and the Provable.*

Cohen, L.J. 1981. Can human irrationality be experimentally demonstrated? *Behavioural and Brain Sciences*, 4:317.

Cohen, N. 1985. Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge. *New York University Law Review*, 60:385.

Cohen, N. 1987. Conceptualizing Proof And Calculating Probabilities: A Response to Professor Kaye. *Cornell Law Review*, 73:78.

Dant, M. 1988 Gambling on the Truth: The Use of Purely Statistical Evidence As A Basis Civil Liability *Columbia Journal of Law and Social Problems*, 22:31.

Eggleston R. 1983. *Evidence, Proof and Probability.* 2nd ed.

Hendrickx, L., Vlek, C. and Oppewal, H. 1989. Relative Importance of Scenario Information and Frequency Information in the Judgment of Risk. *Acta Psychologica*, 72:41.

Kahneman, D. and Tversky, A. 1982a On the psychology of prediction. In Kahneman, D., Slovic, P. and Tversky, A. (ed.), *Judgment under uncertainty: Heuristics and biases*, 48.

Kahneman, D. and Tversky, A. 1982b The simulation heuristic. In Kahneman, D., Slovic, P. and Tversky, A. (ed.), *Judgment under uncertainty: Heuristics and biases*, 201.

Kaye, D. 1979. The Paradox of the Gatecrasher and Other Stories, *Arizona State Law Journal*, 101.

Koehler and Shaviro 1990, Verticial Verdicts: Increasing Verdict Accuracy Through the Use of Overtly Probabilistic Evidence and Methods, *Cornell Law Review*, 75:247.

La Rue, L.H. 1992. Stories versus Theories at the Cardozo Evidence Conference: It's just another metaphor to me. *Cardozo Law Review*, 14:121.

Lempert, R. 1988. The New Evidence Scholarship. In Tillers, P. and Green, E. (ed.) *Probability and Inference in the Law of Evidence: The Uses and Limits of Bayesianism*, 61.

Loftus, E.F. 1979, *Eyewitness Testimony.*

Mackie, J.L. 1981. Propensity, evidence and diagnosis. *Behavioural and Brain Sciences*, 4:345.

Pennington, N. and Hastie, R. 1991. A Cognitive Theory of Juror Decision Making: The Story Model. *Cardozo Law Review*, 519.

Pearl, J. 1988. *Probabilistic reasoning in intelligent systems: Networks of plausible inference.*

Saks, M. and Kidd, R. 1980. Human Information Processing and Adjudication: Trial by Heuristics. *Law and Society Review*, 15:123.

Shaviro D. 1991, A Response to Professor Allen, *Tulane Law Review*, 65:1111.

Simon, R.J. and Mahan, L. 1971. Quantifying burdens of proof. *Law and Society Review*, 5:319.

Slovic, P., Fischoff, B. and Lichtenstein, S. 1976. Cognitive Processes and Societal Risk Taking. In Carroll, J. and Payne, J. (ed.) *Cognition and Social Behaviour*, 172.

Tribe, L.H. 1971. Trial by Mathematics: Precision and Ritual in the Legal Process. *Harvard Law Review*, 84:1329.

Tversky, A. and Kahneman, D. 1982a. Judgment under uncertainty: Heuristics and biases. In Kahneman, D., Slovic, P. and Tversky, A. (ed.), *Judgment under uncertainty: Heuristics and biases*, 3.

Tversky, A. and Kahneman, D. 1982b. Evidential impact of base rates. In Kahneman, D., Slovic, P. and Tversky, A. (ed.), *Judgment under uncertainty: Heuristics and biases*, 153.

Tversky, A. and Kahneman, D. 1982c. Availability: A heuristic for judging frequency and probability. In Kahneman, D., Slovic, P. and Tversky, A. (ed.), *Judgment under uncertainty: Heuristics and biases*, 163.