SIXTH WILFRED FULLAGAR MEMORIAL LECTURE
"BEYOND REASONABLE DOUBT"

SIR RICHARD EGGLESTON*

Just over forty years ago, I was lucky enough to hold a junior brief to the great jurist whose memory we are honouring tonight. The case was *Martin* v. *Osborne,* and I am happy to say that justice triumphed. It was an important decision in the law of evidence, which has occupied much of my attention in recent years, and I refer to it for the purpose of quoting from the judgment of Mr Justice Dixon, as he then was. He said

> "If an issue is to be proved by circumstantial evidence, facts subsidiary to or connected with the main fact must be established from which the conclusion follows as a rational inference. In the inculpation of an accused person the evidentiary circumstances must bear no other reasonable explanation. This means that, according to the common course of human affairs, the degree of probability that the occurrence of the facts proved would be accompanied by the fact to be proved is so high that the contrary cannot reasonably be supposed."[1]

This passage emphasises that the proof of a criminal charge requires a high degree of probability, and this evening I want to examine various questions that arise in relation to this standard of proof.

The same idea had been expressed more than a century before by the great French philosopher, the Marquis de Laplace:

> "In order to condemn an accused person it is necessary without doubt that the judges should have the strongest proofs of his offence. But a moral proof is never more than a probability; and experience has only too clearly shown the errors of which criminal judgments, even those which appear to be the most just, are still susceptible. The impossibility of amending these errors is the strongest argument of the philosophers who have wished to proscribe the penalty of death. We should then be obliged to abstain from judging if it were necessary for us to await mathematical evidence. But the judgment is required by the danger which would result from the impunity of the crime."[2]

In other words, if we wait until we have "mathematical certainty", no criminals will ever be punished. We must therefore steer a course between

---

[1] (1936) 55 C.L.R. 367, 375.
[2] Laplace, *A Philosophical Essay on Probabilities* Tr. by Truscott and Emory, (New York, Dover Publications, 1814), 133.

punishing too many innocent persons (for in the absence of "mathematical certainty", some mistakes are bound to be made) and letting too many guilty persons go free. The problem is to define in some way an acceptable level of probability which will reduce the first type of error to a reasonable level, while not increasing the probability of the second type of error to a point where, as Laplace put it, too great a danger would result from the impunity of the crime.

English lawyers, characteristically, have shrugged this burden off by transferring it to the shoulders of the jury, and they have done so by the use of their favourite device, the test of reasonableness. In the ordinary criminal case, the jury is directed by the judge that the burden of proof (with certain exceptions which I need not elaborate here) is on the prosecution, and that the jury must not convict unless it is satisfied beyond reasonable doubt that the accused is guilty. In cases of circumstantial evidence, it is sometimes considered appropriate to give a further direction of the kind illustrated by the quotation from Sir Owen Dixon in *Martin* v. *Osborne,* but such a direction is quite inappropriate where the task of the jury is to decide between conflicting versions of the facts, as the High Court of Australia pointed out in the recent case of *La Fontaine* v. *R.*[3]

The direction in cases of circumstantial evidence is sometimes expressed in the form that the jury must be satisfied that there is no "rational" hypothesis consistent with innocence, and some judges have spoken as if "rational" and "reasonable" were interchangeable in this context.[4] But, as Barwick C.J. said in *La Fontaine's* case

"A rational conclusion and a rational explanation cannot be equated in the administration of the criminal law with a reasonable conclusion and a reasonable explanation. The jury set for themselves the perimeters of what is, in these contexts, reasonable."[5]

His Honour was there discussing the inappropriateness of the circumstantial evidence test in a case of direct testimony, but the observation is, I think, of general application. A hypothesis may be rational in the sense that it would not offend common sense to find that it had occurred, while at the same time it might be so improbable a hypothesis that it should be rejected. For this reason I prefer the word "reasonable", as in the formulation of Sir Owen Dixon, quoted above, that "according to the common course of human affairs the degree of probability that the occurrence of the facts proved would be accompanied by the fact to be proved is so high that the contrary cannot reasonably be supposed".

The normal formula "satisfied beyond reasonable doubt" has never been seriously questioned in Australia. It has often been said that the

---

[3] (1976) 11 A.L.R. 507, 514, 521, 524.
[4] See Gibbs J. in *La Fontaine's* case, 521.
[5] Ibid. 514.

words are self-explanatory, and that it is dangerous to attempt to elaborate them. Despite these warnings, judges have sometimes tried to help the jury to understand what doubts are to be considered reasonable.

In one case, the judge told the jury to

"consider [the case] in an ordinary common sense manner and in the way you would consider the more serious matters which come up for consideration and decision in your lives, and if in considering it in that way you come to the conclusion—you come to a feeling of comfortable satisfaction that the accused is guilty, then you should find him so guilty . . ."

The High Court of Australia unanimously held that the verdict could not stand, Sir Wilfred Fullagar remarking that the direction just quoted "tends to water down and qualify the plain rule that what is required to justify a conviction is proof beyond reasonable doubt".[6]

It will be recalled that Descartes said that when he tried to see what things he could doubt, he could be certain of nothing except his ability to doubt. He said

"I saw that I could conceive that I had no body, and that there was no world or place where I might be; but yet that I could not for all that conceive that I was not. On the contrary, I saw from the very fact that I thought of doubting the truth of other things, it very evidently and certainly followed that I was . . ."[7]

One might perhaps explain the concept of reasonable doubt by contrast with the philosophic doubt of which Descartes was speaking, but it would be dangerous to assume that members of the jury were equipped with a knowledge of the Cartesian philosophy.

The most that one can safely say is that the doubts to which the judge should pay attention must not be merely fanciful. Wise trial judges go no further than this, if indeed they are prepared to go so far. As Sir Owen Dixon said in *Dawson* v. *The Queen*[8]

"it is a mistake to depart from the time honoured formula. It is, I think, used by ordinary people and is understood well enough by the average man in the community. The attempts to substitute other expressions, of which there have been many examples not only here but in England, have never prospered."

Unfortunately, the English courts have not been able to resist the temptation to try to improve on the time-honoured formula. Lord Goddard, in the early 1950s, seems to have been assailed by doubts as to the ability of juries to understand the expression "beyond reasonable

---

[6] *Thomas* v. *The Queen* (1960) 102 C.L.R. 584, 593; see also *Green* v. *The Queen* (1971) 126 C.L.R. 28.
[7] Descartes, *Discourse on Method (Philosophical Works* tr. by Haldane and Ross, Vol. I) 101.
[8] (1961) 106 C.L.R. 1, 18.

doubt" and recommended that juries be told that before convicting the accused they "must be satisfied of the prisoner's guilt and feel sure of it".[9] Presumably in response to the encouragement of the Lord Chief Justice, judges began to invent new forms of charge, many of which met with disapproval in the Court of Criminal Appeal. By 1960 the traditional phrase seemed to be back in favour,[10] but the alternatives suggested by Lord Goddard were not forbidden. 1961 was a particularly bad year for trial judges. In no less than seven cases, the Court of Criminal Appeal expressed disapproval of the trial judge's charge, though in two of those cases the offending formula was held not to be misleading in the light of the other parts of the summing up. Among the expressions disapproved were "reasonably sure", "pretty certain", and "pretty sure". In one case the conviction was quashed because there was no reference to the words "satisfy", "sure" or "reasonable doubt", and in two cases the judge had used the word "satisfied", or the phrase "proved to your satisfaction", without indicating the degree of satisfaction required. A writer in the Criminal Law Review acidly remarked that it was surely not too much to expect that trial judges should learn the accepted formula and so avoid this waste of resources. The matter was particularly serious in England as there was no power to order a new trial, so that if a mistake was made accused persons who might well have been convicted if a proper direction had been given would go free.[11] Nevertheless, there has been no attempt to standardise the formula, and in *Henry Walters* v. *The Queen*,[12] Lord Diplock for the Privy Council quoted with approval a statement of Lord Goddard

> "If the jury are made to understand that they have to be satisfied and must not return a verdict against a defendant unless they feel sure, and that the onus is all the time on the prosecution and not on the defence, then whether the judge uses one form of language or another is neither here nor there."[13]

Even more unfortunately, His Lordship also gave approval to part of the trial judge's direction in which he attempted to explain what is meant by a reasonable doubt. The trial judge had said "a reasonable doubt is that quality and kind of doubt which, when you are dealing with matters of importance in your own affairs, you allow to influence you one way or the other". Lord Diplock said that such a direction was "unexceptionable", but it is clear that jurors in the conduct of their own affairs, even in matters of importance, may often have to ignore substantial doubts about the correctness of what they are doing. To imply that in the

---

[9] *Practice Note (Reasonable Doubt)* [1952] 1 T.L.R. 1164. A different formulation is given in the report of the same case in [1952] 1 All E.R. 1059.
[10] See [1960] *Crim.L.R.* 630.
[11] See comment by J. C. Smith in [1961] Crim.L.R. 282, 283.
[12] [1969] 2 A.C. 26.
[13] *R.* v. *Kritz* [1950] 1 K.B. 82, 89.

exercise of their functions as jurors in a criminal case they should adopt the same standard is seriously to weaken the protection of the accused, as was pointed out by Sir Edmund Barton in 1913,[14] and again by Sir Wilfred Fullagar in *Thomas* v. *The Queen,* quoted above. Indeed, the danger of such a direction had been pointed out over a hundred years ago, in the notes to the report of *Reg.* v. *White*.[15]

It is clear from Sir Owen Dixon's formulation of the proposition relating to circumstantial evidence that the standard in criminal cases, however it is formulated, requires a high degree of probability. This does not imply, however, that the degree of probability is the same for every class of case. In *Bater* v. *Bater*,[16] Lord Denning had said

> "In criminal cases the charge must be proved beyond reasonable doubt, but there may be degrees of proof within that standard. As Best C.J., and many other great judges have said, 'in proportion as the crime is enormous, so ought the proof to be clear'. So also in civil cases, the case may be proved by a preponderance of probability, but there may be degrees of probability within that standard."

I have not been able to trace the citation from Best C.J. to which Lord Denning referred, but in the case of *Sarah Hobson,* Holroyd J. said "[t]he greater the crime the stronger is the proof required for the purpose of conviction".[17] Lord Denning's statement was quoted with approval by Lord Pearce in the House of Lords, and by Hodson L.J. in the Court of Appeal.[18] Moreover, this view of the matter accords with the views of judges and juries, according to research done in the United States. The following table represents the result of a questionnaire administered to judges, jurors and sociology students by Rita James Simon and Linda Mahan of the University of Illinois.[19]

TABLE 1
PROBABILITY ESTIMATED BY CRIME

Crime	Judges	Mean of Persons Surveyed Jurors	Students
Murder	.92	.86	.93
Forcible Rape	.91	.75	.89
Burglary	.89	.79	.86
Assault	.88	.75	.85
Petty Larceny	.87	.74	.82

[14] *Brown* v. *The King* (1913) 17 C.L.R. 570, 586.
[15] (1865) 4 F. & F. 383; 176 E.R. 611.
[16] [1951] P. 35, 36, 37.
[17] (1823) 1 Lewin 261; 168 E.R. 1033. See also the reference to *Re Bland* in the notes to *Reg.* v. *White* cited above.
[18] See *Blyth* v. *Blyth* [1966] A.C. 643, 673; *Hornal* v. *Neuberger Products Ltd.* [1957] 1 Q.B. 247, 263, 264.
[19] Rita James Simon and Linda Mahan, "Quantifying Burdens of Proof" (1971) 5 *Law and Society Review* 319.

The authors asked the subjects of their research to specify in relation to various crimes the degree of probability that they would require in order to achieve satisfaction beyond reasonable doubt. I have simplified the table somewhat, by omitting a number of the offences, and I have arranged them in what I would regard as descending order of seriousness. My order happens to correspond with the views of the judges and the students, but the jurors, who were selected, with the co-operation of the Chief Judge of the Champaign County Court, from persons serving on their regular period of jury service, have a slightly different list of priorities.

I have also converted the probability evaluations into decimal form instead of the scale ranging from 1 to 10 used by the authors. There are several different ways in which one can express one's evaluation of a probability. If we draw a card from a normal well-shuffled pack, there are thirteen chances in 52, or one chance in four, that it will be a spade. We can express this as a probability of one-quarter, 0.25 or 25%, or we can say the odds are 3 to 1 against. Scientists usually express probabilities in decimal form, on a scale ranging from zero (for events which are considered to be impossible) to 1, for events considered to be certain.

It is of course impossible in most cases in the courts to arrive at a figure which represents the probability of guilt, or the probability, in a civil case, that the defendant is liable. There are cases in which some part of the evidence can be assigned a numerical figure representing its probability evaluation. For example, we can often ascertain from statistics compiled by scientists what is the probability that a particular person will belong to a particular blood group. But cases in which the ultimate question of fact can be evaluated in this way must surely be extremely rare. Nevertheless, it is possible to get some idea of the strength of a person's opinion as to the likelihood of an event by asking him to make an evaluation in mathematical form of the probability, even though no mathematician could arrive at a figure by any quantitative means.

Table 1 gives an average figure for each group studied (judges, jurors and students) for the various types of crime. But obviously this figure is subject to considerable variation as between individual judges or jurors or students. Table 2 is a simplified table taken from the same source. Each subject was asked a general question (not referring to any specific type of crime) as to the probability he or she would require before convicting. The table shows how the answers were distributed within each class.

The upper part of the table is to be read in the following way: taking the judges, 4% of those surveyed were prepared to find a man guilty of a criminal offence (applying the "beyond reasonable doubt" standard) where the probability of guilt as they evaluated it did not reach 0.7. One third would require a level between 0.7 and 0.9, while 63% of them would require a probability of at least 0.9 before finding a man guilty. It

will be noted that in the juror group there were many more in the lower categories, while students were much closer to the judges in their standards.

TABLE 2
LEVEL OF PROBABILITY REQUIRED FOR "GUILTY" VERDICT

Probability	Judges	Jurors	Students
"Beyond Reasonable Doubt" Standard			
Below .7	4%	26%	7%
.7 and below .9	33%	20%	28%
Above .9	63%	54%	65%
"Preponderance of Evidence" Standard			
Below .7	81%	22%	21%
.7 and below .9	13%	46%	51%
Above .9	6%	32%	28%

In an investigation conducted by Miss Bridget Walsh, a student of Professor John Cohen of the Department of Psychology of Manchester University, it was found that one third of the subjects fixed a level below 0.7, one third thought the level should be above 0.9, while the remaining third chose a level between those figures.[20] The figures given by Simon and Mahan (not reproduced in full in Table 2) would suggest that the lowest third of the jurors would require less than 0.75, the middle third from 0.75 to 0.95, and the top third more than 0.95.

The discrepancy shown in Table 2 between the standards of judges and juries might be taken to mean that juries would be more ready to convict than judges. An American investigation by Kalven and Zeisel, however, reveals that in that country juries tend to be more lenient, i.e. to acquit more often than judges would.[21] The authors sent out a questionnaire to some 3500 judges, of whom 555 responded by filling in forms covering 3576 trials, and recording whether they agreed or disagreed with the verdict of the jury. Broadly speaking, the returns showed that the judge was more lenient than the jury in 3% of cases, and the jury more lenient than the judge in 19% of cases. There does not seem to be any similar comparison available for England, but the authors of the American study asked Lord Parker of Waddington, then Lord Chief Justice, about the English experience. Lord Parker consulted all the senior Queen's Bench judges and replied that none of them had kept any figures, but from their general impressions they felt that it was rare for a judge to feel that he would have acquitted where the jury in fact convicted. Cases where the jury acquitted but the judge would have convicted were more frequent, and estimates ranged from 3% to 10% of acquittals. At the same time,

[20] See Cohen and Christensen, *Information and Choice* (Edinburgh, Oliver and Boyd, 1970) 62, 63.
[21] Kalven and Zeisel, *The American Jury* (University of Chicago Press, 1971) 59.

cases in which the judge considered the verdict perverse were also very rare: that is to say, in those cases in which the judge disagreed with the jury's acquittal he was usually prepared to concede that the jury may have been right.[22]

There are, of course, many reasons why juries might be expected to be more lenient than judges, other than differences in the standard of proof, but Kalven and Zeisel, after eliminating as many of those reasons as they could identify (such as the judge's knowledge of facts not known to the jury) still found that there was a residue of acquittals unexplained. On the assumption that the standard of persuasion required by a juror would not differ significantly from that of a judge, they explained the difference by reference to the requirement that the jury should be unanimous. The standard applied in a trial, they thought, would tend to be, not the average of the twelve, but that of the jurors who insist on certainty or near-certainty.[23] Of the 69 jurors tested by Simon and Mahan, 25 opted for a probability of 1.0, i.e. absolute certainty. From these figures it seems likely that most juries would contain three or four members whose standard would be certainty or very close to it.

It should be added, in fairness to the American figures, that in that country there is usually an option to waive the right to trial by jury, and it seems that this right is freely availed of where it is thought that the jury is unlikely to be more favourable than a judge; but where it is considered that a jury is likely to be more lenient than a judge, the accused elects for trial by jury. Kalven and Zeisel's research revealed that the Bar are remarkably good at forecasting which classes of crime are in that category. In England, the options are not nearly as wide, and in general, if the offence is indictable, it must come before a jury unless the accused pleads guilty.

The lower part of Table 2 summarises the answers given by the groups surveyed by Simon and Mahan when they were asked to apply the "preponderance of evidence" standard. This is the standard of proof for civil cases in the United States, and it corresponds in essence with the civil standard in England and Australia.[24] Looking at this table, it will be seen that the trend shown in the upper part is reversed: 81% of judges would be content with a level lower than 0.7, while only 22% of jurors would accept so low a standard. It is possible, however, that the results are affected by the circumstances under which the opinions were obtained. The same questionnaire was given to judges, jurors and students, but the jurors and students received the questionnaire after they had listened to a tape recording of a trial for homicide, and there may have been some

---

[22] These conclusions were for cases of "ordinary crime", excluding cases of causing death by dangerous driving: Kalven and Zeisel, op. cit. 514, 515.

[23] Kalven and Zeisel, op. cit. 189.

[24] See Simon and Mahan, op. cit. note 2.

residual effect of the frame of mind induced by being asked to find the accused guilty or not guilty of murder.

In England, a team from the London School of Economics has undertaken a project designed to test the effect of different kinds of evidence, and of different standards of proof as embodied in the summing up. The research did not, as the American project did, provide a comparison between the standards of judges and the standards of juries. Nevertheless, comparison of the views of the jury as to "beyond reasonable doubt" and "more likely than not" suggested the same tendency for the jury to set a higher standard for "more likely than not" than one would have expected. More significantly, the experiments disclosed a significant difference between "beyond reasonable doubt" and "sure and certain". The figures so far as relevant for our purposes are set out in Table 3.

TABLE 3

NUMBERS AND PERCENTAGES FOR GUILTY (G) AND NOT GUILTY (NG)

Cases	Beyond Reasonable Doubt		Sure and Certain		More Likely than Not	
	G	NG	G	NG	G	NG
Theft Number	43	94	32	58	42	50
% G	31		35		46	
Rape 1st Defendant Number	94†	48	39†	32	30†	14
% G	66		55		68	
Rape 2nd Defendant Number	46†	96	13†	58	18†	26
% G	32		18		41	
All Cases Number	183	238	84	148	90	90
% G	43		36		50	

† Includes "Guilty of Attempted Rape".

The trials to which the jurors listened were slightly shortened versions of actual trials, recorded on tape by actors. Different juries heard slightly differing versions of the evidence, and the summing up varied in several respects, the important variation for our purposes being the variation in the direction as to the standard of proof.

The two cases chosen for the experiment were a theft case, in which the accused was charged with having stolen a quantity of meat, and a rape case in which two defendants were involved. Each juror was asked to record his verdict at the conclusion of the summing up, and again after the jury had discussed the case. No joint verdict was taken, the jury being treated as "hung".

Two rather unexpected points of interest emerged, though they are not directly relevant for our present purposes. The first was this: in some

of the trials, evidence of prior convictions was introduced, sometimes accompanied by a warning from the judge to disregard the evidence, and sometimes without. In some cases the prior convictions were for similar offences to that charged (e.g. previous convictions for theft in the theft trial) in others for offences of a different kind (e.g. previous convictions for dishonesty in the rape trial). When the prior convictions were for offences of a similar type, and no warning was given that they were to be disregarded, the rate of conviction was greater, as one would expect, but when they were for offences of a different type, the conviction rate not only did not increase, it actually went down, as if the jury had reacted to what it regarded as an improper attempt to influence it by reference to the prior convictions.

The second point of interest was that in the rape case, the effect of introducing the second defendant's prior convictions for similar offences was to increase the frequency of convictions against the first defendant even though he had no such previous record. On the other hand, when the jurors were instructed to disregard the prior convictions, the rate of conviction tended to drop, indicating that juries do in fact take some notice of what judges tell them in such cases.

However, this is a digression. Our concern is with standards of proof, and it is for this purpose that I have extracted the figures in Table 3.

What the table shows is that on the whole, a direction to the jury that they must be "sure and certain" tends to reduce the rate of convictions (as compared with "beyond reasonable doubt") while a direction based on the civil standard ("more likely than not") tends to increase the rate. The increase in the rate brought about by "more likely than not", however, is not as great as one would expect. In particular, in the case of the first defendant in the rape case there is hardly any difference between "beyond reasonable doubt" and "more likely than not". Of course, comparison of these results in detail can be misleading. Where there is a very strong case against the accused, the effect of the difference between the two forms of summing up will probably be less than if the case is weak. To take the extreme case, if the evidence is overwhelming, the rate of conviction would be 100% whichever form of instruction was given to the jury. Where the case is closer to the borderline, "more likely than not" might produce a substantial number of convictions, but "beyond reasonable doubt" hardly any.

What is more significant for our purposes is the difference between the conviction rate for "beyond reasonable doubt" and that for "sure and certain". It will be seen that in the rape case the conviction rate for both defendants, but especially the second, is lower for "sure and certain" than for "beyond reasonable doubt". In order to appreciate the difference in the two forms of direction as heard by the "jury", the text of the actual directions may be helpful. They are as follows:

In the theft case, for "beyond reasonable doubt"

"You should be sure beyond reasonable doubt and by a reasonable doubt I mean not a fanciful doubt, but such a doubt that might affect you in daily business or domestic decisions."

For "sure and certain"

"Before you convict you must feel sure and certain on the evidence you have heard that the accused is guilty."

In the rape case, for "beyond reasonable doubt"

"You should be sure beyond reasonable doubt and by reasonable doubt I mean not a fanciful doubt that you might use to avoid an unpleasant decision, but a doubt for which reasons can be given."

For "sure and certain"

"Before you convict you must feel sure and certain on the evidence you have heard that the accused is guilty" (this is the same formula as used in the theft case).

The point about the figures we have been looking at is that they seem to confirm the conclusion arrived at in the American study and the Manchester research, that for a significant number of people, "beyond reasonable doubt" means something less than certainty. I doubt whether Lord Goddard, when he recommended the adoption of the "sure and certain" type of formula, thought that he might be making it easier for criminals to escape conviction. It would seem that if we wanted to raise the standard of proof above the present level, we could consider the adoption of Lord Goddard's formula.

Before we place too much reliance on figures of this kind, however, let me draw attention to the results of some other research, which indicates that the verdicts of juries are not based entirely on rational considerations. It seems that the order in which the evidence is presented to a jury may make a considerable difference to the result of a trial. In an English experiment, the evidence in a bigamy trial was presented to one group of fifty-six subjects in a particular order. To another group of fifty-six the same evidence was presented, but in a different order. In the first group, only three out of fifty-six were for "Guilty"; in the second group, eighteen out of fifty-six were prepared to find the accused guilty.[25]

A somewhat similar phenomenon was noticed by Simon and Mahan in the Illinois research. Here, the difference was not in the order of presentation of the evidence, but in the order in which questions were asked of the jury after they had listened to the recorded evidence. There was found to be a marked difference in the percentage of guilty verdicts, according to whether the subjects were asked to say whether the accused was guilty or not guilty before they were asked to evaluate the probability in mathematical terms, or whether they were asked for a probability

---

[25] Audley, "What Makes Up a Mind?" in *Decisions, Organizations and Society,* (London, Penguin, 1971) 62, 63.

estimate first and then asked whether they thought the accused was guilty or not guilty. The results are set out in Table 4.

TABLE 4
PERCENTAGES FOR GUILTY

		Jurors	*Students*
Before Deliberation			
Group A	Verdict ▶ Probability	51%	20%
Group B	Probability ▶ Verdict	38%	9%
After Deliberation			
Group A	Verdict ▶ Probability	32%	16%
Group B	Probability ▶ Verdict	19%	5%

Here the jurors were divided into two groups, which I have labelled Group A and Group B. Each juror in each group was asked to say whether the accused was guilty or not guilty, but those in Group A were asked that question first, and then asked to give a probability estimate on a scale from one to ten. After about thirty minutes' discussion (for which purpose they were divided into groups of six), they were again asked the same questions, and in the same order. In Group B, the subjects were asked to give a probability estimate first, and then to say whether the accused was guilty or not guilty. After discussion in groups of six, the same questions were asked again.

Two things will be noticed about the table. In the first place, the effect of discussion with their fellows appears to have been to reduce the number who thought the accused was guilty. The second thing is that asking a juror to make a probability estimate before he gives his verdict "Guilty" or "Not Guilty", reduces the number of guilty verdicts.[26]

This second effect, however, must be interpreted with caution. Experiments have shown that people who are asked to estimate probabilities tend to underestimate the numerical evaluation, at least when the probability is high. For example, subjects are shown two bags, identical in appearance. They are told that one contains 70 green poker chips, and 30 white; the other 70 white chips and 30 green. Twelve chips are drawn from one of the bags and turn out to be 8 green and 4 white. What is the probability that it is the predominantly green bag (70 green to 30 white) from which the chips were drawn? Assuming that the bag is chosen at random and the chips are drawn at random, mathematicians would compute the probability that the bag is the predominantly green bag at 0.977, or 97.7% but the subjects of the experiment evaluate the prob-

---

[26] I have assumed, though it is not specifically stated by the authors, that the composition of Groups A and B was the same throughout, i.e. that those who were asked the question in a particular order before the group discussion were asked the questions in the same order after the discussion.

ability at from 0.6 to 0.8. Conversely, subjects tend to overestimate the probability when it is very low.[27] A somewhat less complicated probability calculation was involved in a bead-guessing experiment conducted by Cohen and Hansel. The subjects were asked (amongst other things) to give a realistic estimate of their chances of drawing blue beads from an urn. They were told to make the estimate in respect of three urns, one containing 90 blue beads and 10 yellow, the second containing 50 blue beads and 50 yellow, the third containing 10 blue beads and 90 yellow. The average estimate for the first urn (with a 90 per cent probability of blue) was 0.78 (78%), for the second (with a 50 per cent probability) 0.52 (52%), and for the third (with a 10 per cent probability) 0.22 (22%). It will be seen that the underestimate of the high probability was exactly matched by the overestimate of the low probability.[28]

If we can assume that the same sort of understatement applies to the evaluations of probability which were asked for in the experiments in the Illinois research, we might explain the shift from guilty to not guilty by saying that some of those who said "not guilty" after making a probability estimate were misled by the fact that, when asked to put a figure on the probability, they mistakenly chose one that underestimated their real view of the strength of the case for the prosecution. Those concerned were presumably those whose numerical estimates of probability were above 0.5 but not far enough above to satisfy their idea of proof beyond reasonable doubt.

Another set of figures may cause one to wonder whether too much should be made of variations found in relatively small samples of experimental work. In 1966, the New Law Journal[29] published figures for the rate of acquittal of defendants who pleaded not guilty in England and Wales.

TABLE 5

RATE OF ACQUITTAL—SELECTED CITIES AND COUNTIES—1966

Area	Tried	Acquitted	%
Monmouthshire	23	21	91
Kent	218	127	58
Manchester	358	162	45
Metropolitan P.D.	1148	448	39
Birmingham	241	85	35
Yorks (W. Riding)	141	46	33
Liverpool	501	139	28
Worcestershire	107	14	13
England and Wales	7765	3029	39

[27] The figure of 0.977 is for drawing without replacement; with replacement it is 0.967; see Fairley, "Probabilistic Analysis of Identification Evidence" (1973) 2 *Journal of Legal Studies* 494-6.

[28] Cohen and Christensen, *Information and Choice* (Edinburgh, Oliver and Boyd, 1970) 88. These were subjects aged 15-16.

[29] Vol. 116, p. 928.

For the year in question, 7765 persons were tried by jury, and 3029 or 39% were acquitted. It so happened that the rate of acquittal for the Metropolitan Police District was also 39%, but in other parts of the country there were wide variations. The figures for Monmouthshire look startling at first sight, as they show that Monmouthshire juries acquitted 91% of the defendants tried in that county. The then Attorney-General, who was a Welshman himself, was asked about the statistics, and he said that Welsh juries were generally in favour of justice, but they were not bigoted about it.[30] In fact, the Monmouthshire figure is for the county excluding Newport. There were 38 trials in Newport and only 23 in the rest of the county. Newport juries only acquitted 29% of defendants, so that the overall rate for Monmouthshire came out at slightly over 50%. But even so, there are wide variations which are difficult to explain. The County of Kent, with 218 trials, acquitted 58%, while Worcestershire, with 107 trials, only acquitted 13%. In Liverpool, with 501 trials, only 28% were acquitted, whereas in Manchester 45% of the 358 defendants were found not guilty. Of course these differences may reflect a variety of factors other than differing views of probability levels—tougher or more lenient attitudes on the part of the police, or on the part of the magistrates who are asked to commit for trial; or differences in the kinds of crime that are prevalent in different areas.

It would seem that the acquittal rate has risen since these figures were compiled. In 1973 Sir Robert Mark, delivering the Dimbleby Lecture for that year, said that the acquittal rate for England and Wales had risen from the 39% given above for 1966, to 50%.[30a] He attributed this increase to the fact that after 1967 cases were no longer effectively sifted by a magistrate, as a result of the introduction of the system of "hand-up briefs" enabling a person to be committed for trial without the witnesses being called to give evidence in the lower court. If this is the true explanation, it would seem that the new procedure has saved time and money at the lower court level, at the expense of an increase in the number of unnecessary trials at the higher level. But I think most practitioners in this country would agree that it is comparatively rare for magistrates to decline to commit an accused person for trial if there is evidence against him, even though it be weak, and if they do commit on very weak evidence, the prosecutor is likely to enter a *nolle prosequi,* so that the accused will not be called upon to plead in the higher court. I think it unlikely, therefore, that the reason given by Sir Robert Mark

[30] See Sir Robert Mark, *Minority Verdict* (London B.B.C., 1973) 10.
[30a] Since this lecture was prepared, my attention has been drawn to data which give rise to some doubt as to whether the increase referred to has actually occurred; see M. Zander, "Are Too Many Professional Criminals Avoiding Conviction?" (1974) 37 *Mod.L.R.* 28, 32; S. J. Elgrod and J. D. M. Lew, "Acquittals—A Statistical Exercise" (1973) 123 *New Law J.* 1104; S. J. Elgrod and J. D. M. Lew, "Acquittals—Further Statistics" (1975) 125 *New Law J.* 45, 46; S. J. Elgrod, "Acquittals—A Further Survey" (1976) 126 *New Law J.* 580.

could account for the increase in the rate of acquittals in England. It may reflect in part a decline in the confidence of juries in the police force, and a greater willingness to believe that a confession alleged to have been made by the accused was not made or was induced by improper means and hence not to be relied on.

It may be worth remarking that during the period of this decline in the success rate for prosecutions, provision was made by statute in England for majority verdicts.[31] One would have expected this provision to reduce the number of acquittals. In Scotland the rate of acquittal is only 20%. They have had majority verdicts for a long time—a simple majority in a jury of fifteen, as compared with the English provision for 10 out of 12, 10 out of 11, or 9 out of 10.[32]

. Attempts have been made from time to time to propose the introduction of more scientific methods into the assessment of guilt, by trying to devise methods of evaluating the probabilities applicable in specific situations. An examination of these suggestions, however, reveals the difficulty of the operation, even in cases where the basic probabilities are capable of more or less exact estimation. Let us suppose, for example, that a partial fingerprint has been left at the scene of a crime. We know that the odds against two persons having the same fingerprint are high, because examination of thousands of prints has revealed that the different kinds of patterns found in fingerprints occur independently and in a very large variety of combinations. Because of this random character, it has been calculated that the chance of two prints taken at random coinciding in all their aspects is less than one in 64,000 million. Actually it is very difficult to find an up-to-date figure for this probability, but that figure was given by Sir Francis Galton many years ago and is generally regarded as conservative.[33] But what does this mean? It does not mean that if we take a particular person's prints, it is 64,000 million to one that no other print can be found with the same characteristics. The figure given was for the chance of duplication in a random selection. We may illustrate the point by an analogy. If we draw a single card from a normal well-shuffled pack, the chance of it being the Ace of Spades is one in 52, or 51 to 1 against. But if we search through the pack until we find an Ace, the chance that it will be the Ace of Spades is one in four, or three to one against. Let us suppose, then, that we find, not a complete print, but a partial print, and for the sake of argument, let us suppose that the experts can tell us that the probability of the duplication of the characteristics of that partial print in a person chosen by chance is one in 1000, or 0.001. If we search a community of 4000 people, we may expect to find four persons in that

[31] *Criminal Justice Act* 1967, s. 13.
[32] Lord Kilbrandon, *Other People's Law* (London, Stevens and Sons, 1966) 97.
[33] Quoted by Kingston, "Probability Theory in Criminalistics" (1965) *Am. Stat. Assn. Journal* 75.

community with the given characteristics, and if there were no other
evidence available, we should not be able to tell which of those four
persons was the guilty one. If we merely searched until we found a person
with the required characteristics, there would be one chance in four that
we had found the guilty person, as in the case of our search of the pack
until we found an Ace. The point about this example is that a situation
has actually occurred in which this sort of argument was relevant. In
California, in the case of *People* v. *Collins*,[34] the evidence was that an old
lady had been robbed by a blonde wearing her hair in a pony tail, who
had fled to a yellow car driven by a negro with a beard and moustache
and been driven away. The prosecutor called a teacher of mathematics
and obtained answers to a series of questions from which he argued that
the odds against finding such a combination of characteristics in a couple
chosen at random were approximately twelve million to one. From this
he argued that the odds were twelve million to one against the accused
couple being innocent, and that he had therefore proved the case beyond
reasonable doubt. The appellate court in California set aside the convic-
tion for several reasons, not least being that the prosecutor and his expert
witness had assumed that each of the characteristics deposed to was likely
to occur independently of the others; for example, it was assumed that the
probability of finding that a person had a moustache was not related to
the probability that he had a beard. But the court also pointed out the
fallacy of the argument which I have just outlined, that if the chances of
duplication in a random selection are low, the probability of identity can
be taken as correspondingly high, irrespective of the population that has
to be searched before you find a match.

The point in this case for our present purposes, however, does not lie
in drawing attention to the fallacies involved in the argument of the
prosecutor, but in the discussion that the case generated amongst math-
ematicians as to the correct way in which to calculate the probability of
identity in such cases. In *Collins'* case, the figure of twelve million
corresponded with the population of the Los Angeles area, where the
crime took place. The Californian Supreme Court calculated that the
probability that there would be a second couple having the same
characteristics would approach 0.43, that is to say, that the probability
that the accused couple were the only such couple in the area would be
only about 0.57 or somewhat less than six to four on (this was assuming
that the prosecutor's figure of one in twelve million was correct). But
the court's mathematics have been questioned by more than one writer,[35]
and at least two suggestions have been put forward as to the correct

[34] (1968) 438 P. 2d 33.
[35] See Finkelstein and Fairley, "A Bayesian Approach to Identification Evidence"
(1970) 83 *Harvard Law Review* 489, 492, 493; Kurt Kreith, "Mathematics, Social
Decisions and the Law" (1976) 7 *International Journal of Mathematical Education
in Science and Technology* 317.

method of calculation.[36] I am not competent to resolve the differences of the mathematicians, but the fact that they exist does not augur well for the application of mathematical theory to the resolution of problems of reasonable doubt.

Another approach to questions of probability was put forward by Finkelstein and Fairley, in the Harvard Law Review.[37] They suggested that it would be possible to make use of Bayes' Theorem in the resolution of problems of guilt, where there was identification evidence in addition to other evidence tending to establish the guilt of the accused. Bayes' Theorem, first enunciated more than two hundred years ago, enables us to modify a pre-existing evaluation of probability in the light of experimental evidence. Thus, to take the example I gave earlier, if we have two bags with different numbers of green and white chips, we choose a bag at random and draw twelve chips (8 green and 4 white), it is Bayes' Theorem that enables us to calculate the probability that the bag is the one with more green chips than white. I need not go into the exact mechanics here. Finkelstein and Fairley illustrate their proposal by an example in which evidence is given in a murder case to the following effect: the accused (A) and the victim (V) are lovers. A is known to have threatened violence to V on previous occasions. V is found stabbed with a knife, and on the handle of the knife is a print which is incomplete, but of which the experts can say that the chance of random duplication is no more than one in a thousand. A's print corresponds with the print found. If we were able to evaluate the probability of A's guilt before we knew about the print on the knife handle, we could use Bayes' Theorem to calculate the new probability in the light of that evidence. Table 6 is taken from their article, and shows the way in which the prior probability is modified by the additional evidence.

TABLE 6
POSTERIOR PROBABILITY P(G/H)

Frequency of Characteristics P(H/NG)	Prior Probability P(G)				
	.01	.1	.25	.5	.75
.5	.019	.181	.400	.666	.857
.25	.038	.307	.571	.800	.923
.1	.091	.526	.769	.909	.967
.01	.502	.917	.970	.990	.996
.001	.909	.991	.997	.9990	.9996

Note: In this Table, P(G/H) represents the probability that the defendant used the knife, assuming that a handprint similar to the defendant's is found; P(G) is the prior

[36] See, e.g., Kingston, "Applications of Probability Theory in Criminalistics" (1965) *Am. Stat. Assn. Journal* 70; Cullison, "Identification by Probabilities and Trial by Arithmetic" (1969) 6 *Houston Law Review* 471; Smith and Charrow, "Upper and Lower Bounds for Probability of Guilt Based on Circumstantial Evidence" 70 *Journal of American Statistical Association 555*.
[37] See fn. 35 supra.

probability, i.e. the probability that the defendant used the knife, considered without regard to the existence of the handprint; P(H/NG) is the probability of finding a handprint of the given description, assuming that the defendant did *not* use the knife. The method of using the table is explained by Finkelstein and Fairley as follows: It is necessary first to estimate the probability that the defendant used the knife ignoring the existence of the handprint (i.e. relying solely on the other evidence). Let us assume that this is evaluated at 0.25. Given that the probability of finding a similar print made by someone other than the defendant is one in a thousand (0.001) we look down the 0.25 column for P(G) and along the 0.001 line for P(H/NG) and find that the probability that the defendant used the knife should now be assessed at 0.997 (see (1970) 83 *Harv. L. Rev.* 498-500).

What Finkelstein and Fairley suggest is that in appropriate cases, the jury should be invited to make a preliminary estimate of probability excluding the evidence about the print, and then be assisted by expert evidence furnishing a range of figures similar to those derived in Table 6 to enable them to work out how they should re-assess the probabilities. Professor Tribe of Harvard University has severely criticised the proposition.[38] He points out, amongst other things, that there are two assumptions involved which may seriously affect the calculation, namely, that whoever committed the murder would certainly have left a print on the knife, and that if the print is in fact A's print, he is certainly guilty. Neither assumption can be justified, and if they are not taken as true, not only is the table inaccurate, but the calculation necessary to allow for the possibility of their being untrue is much more complicated and involves considerable elements of guesswork. But Tribe's objections are not confined to the mathematical details of the process. In particular, he attacks the whole concept of trying to put a mathematical figure on the probability of guilt. He says that the concept of guilt beyond a reasonable doubt

"signifies not any mathematical measure of the precise degree of certitude we require of juries in criminal cases, but a subtle compromise between the knowledge, on the one hand, that we cannot realistically insist on acquittal wherever guilt is less than absolutely certain, and the realization, on the other hand, that the cost of spelling that out explicitly and with calculated precision in the trial itself would be too high."[39]

A more recent attempt to introduce probability theory into crime detection and prosecution is to be found in an article by Lindley,[40] in which he analyses a "simple situation" in which samples of hair corresponding to the hair of the suspect are found at the scene of the crime. He finally arrives at a formula for calculating the probability of guilt, and says

"That part of the analysis that involves only sampling and laboratory errors . . . requires only the necessary statistical expertise. The rest of

---

[38] See Tribe, "Trial by Mathematics" (1971) 84 *Harv. L. Rev.* 1329. The debate was continued in 84 *Harv. L. Rev.* 1801, 1810.
[39] 84 *Harv. L. Rev.* 1375.
[40] "Probabilities and the Law", in *Utility, Probability and Human Decision Making* (Dordrecht, Holland, Reidel Publishing Co., 1975) 223, 229, 230.

the evaluation of *both* sets of probabilities in [the final equation] could perhaps be carried out by a committee composed of lawyers, statisticians, Home Office officials and lay members who could reach a reasoned assessment of the chances involved and could report their values at any relevant trial."

Even with this machinery for the assessment of probabilities, however, Lindley recognises that some probabilities would still be likely to continue to be a matter for the court. Moreover, while he thinks "beyond reasonable doubt" appears to indicate odds of about 100 to 1, he presumes that "the probability required is related to the seriousness of the offence" which, as we have seen, is the view taken by judges and juries.

The same difficulty occurred to the former Director of the Metropolitan Police Laboratory in London, H. J. Walls, in an article written in 1971.[41] After discussing the sort of problems involved in *People* v. *Collins* and in the example postulated by Finkelstein and Fairley, he says

"It appears to me that at present the most useful next step would be to decide what sort of probability figure represents 'beyond reasonable doubt'. This seems to be not an impossible task. Assume that a number of cases could be found in which the evidence admits of a realistic and meaningful calculation of probabilities. The first step would then be to 'try' these cases in parallel—on the one hand by a statistician calculating probabilities, on the other by a court (if the case actually came to trial) or a panel of lawyers (if it did not). A comparison of the results might then enable us to decide what sort of probability figure is equivalent to proof 'beyond reasonable doubt'. If anything approaching consistency emerged from these comparisons, something very useful would have been discovered."

Having got so far, it seems to have occurred to the author that if the degree of probability required for proof beyond reasonable doubt varied according to the seriousness of the crime (as to which, not being a lawyer, he offered no opinion), the experiment would not be nearly as simple as that.[42]

Our examination of the problem has shown that there are other reasons why such an experiment would be unlikely to produce useful results. In the first place, as Lindley points out, even in simple cases, there will still be probabilities which only the tribunal can evaluate. Secondly, the statisticians themselves have not yet agreed on how to calculate the probabilities, even given the assumptions made in the *Collins'* case as to the odds against finding a duplicate with the same characteristics as the criminal. Thirdly, evaluations by judges and juries might give very different results, and if the evaluation is to be made by juries (which would be the more appropriate basis of comparison) from which State or

[41] "What is 'Reasonable Doubt'?" [1971] *Crim.L.R.* 458, 469.
[42] Ibid. 470.

County would you draw your jury, and how many experiments would you need to obtain significant comparisons? Finally, we have seen that numerical evaluations of probabilities by laymen may be very wide of the mark.

The truth is that if we are to attempt to achieve consistency in the evaluation of the probability of guilt, we would have to scrap the jury system and hand over the job to scientists. It could hardly be expected that twelve men drawn at random from a jury panel would be able to handle the sort of statistical material involved in probability calculations, even with the assistance of experts, who might in any event disagree among themselves as to the correct approach. But would scientists be any better at evaluating the unquantifiable elements that enter into almost every criminal case? Recently, for another purpose, I read the trial of Elizabeth Canning, who was prosecuted for perjury in 1754. It appears that during a period of nearly four weeks in 1753 she went missing, and on her return told a remarkable tale of how she had been taken by two ruffians to a house at Enfield Wash, near London, where a woman cut off her stays and then locked her in a room with some bread and water, on which she subsisted until she escaped nearly four weeks later. She subsequently identified the woman as a gypsy named Mary Squires, who was tried for the theft of the stays and found guilty, on Elizabeth Canning's evidence. There were serious discrepancies in the story told by Elizabeth Canning, and Mary Squires was pardoned and in due course Elizabeth was tried for perjury. In order to prove that her story in the previous trial was untrue, thirty-eight witnesses were called to establish that during the period of Elizabeth Canning's absence, Mary Squires and her son and daughter were either in Dorset, or travelling from Dorset to London, which they reached just in time for Mary Squires to be identified by Elizabeth as the woman who had robbed her of her stays. On the other hand, twenty-seven witnesses deposed to having seen Mary Squires at times and in places where she could not possibly have been if the evidence of the prosecution witnesses was correct. I should add that Mary Squires did not have a forgettable face; indeed, she was described during the case as "hideous" and "frightful", and by one witness as having a face unlike any other that the witness had ever seen.

In the event, Elizabeth Canning was convicted of perjury. At first, the jury reported their view that she was guilty of perjury, but that it was not wilful. Since perjury must be "wilful and corrupt", the judge told them that he could not accept such a verdict, and after further deliberation they returned a verdict of guilty without qualification. Before she was sentenced, an attempt was made by two of the jurors to show that they had not agreed in the verdict; they said that they had understood that the words "wilful and corrupt" were mere verbiage, and that they did not believe that she had lied deliberately, but the court declined to accept their "recantation" and sentenced her to be transported to the Colonies for

seven years (this of course meant America, not Australia, in 1754). As Mary Squires had been sentenced to death for the theft of the stays, and had only narrowly escaped the gallows, the sentence can hardly be regarded as too severe, if in fact Elizabeth Canning was guilty.

What would a panel of statisticians, or a committee of lawyers, statisticians, Home Office officials and lay members have made of the conflict between the thirty-eight witnesses on one side and the twenty-seven on the other? The fact is that although writers on probability have been concerned with the problem of credibility for centuries, no solution has yet been found for the problem of computing the probabilities when the reliability or the credibility of witnesses is in question.[43] Accordingly, it seems likely that we shall be relying on human judgment and common sense in such cases for some time yet. Personally, I do not feel distressed at this prospect. Although some of my best friends are statisticians, I would prefer a jury of twelve to a panel of statisticians if ever I were charged with an indictable offence.

The Australian Broadcasting Commission is currently[43a] showing a series of programmes with the same title as this lecture. This is a coincidence, as I had chosen the subject of this lecture last November, and gave it its title some time before the television series was announced. I should observe, however, that the programmes shown so far have been of cases in which it is suggested that despite the conviction some doubt as to the guilt of the accused still remains. I do not propose to comment on this except to link the suggestion up with the theme of my discourse. To say that "beyond reasonable doubt" implies something less than what Laplace called "mathematical certainty" is to admit, what I suppose all lawyers would admit, that mistakes can sometimes be made. Indeed, leaving aside the current television series, there are well-authenticated cases in which juries have convicted innocent men. But, as Professor Tribe pointed out, this is not the same thing as saying that the jury should fix a figure which would represent the number of innocent persons they would be prepared to convict in order to protect society from the guilty. Kalven and Zeisel have said that the requirement of proof beyond reasonable doubt

> "is a way of saying that we live in a society that prefers to let ten guilty men go free rather than risk convicting one innocent man . . . In the end the point is that the jury, as an expression of the community's conscience, interprets this norm more generously and more intensely than does the judge. If a society wishes to be serious about convicting only when the state has been put to proof beyond a reasonable doubt, it would be well advised to have a jury system."[44]

---

[43] See Kingston, "Probability and Legal Proceedings" (1966) 57 *Journal of Criminal Law, Criminology and Police Science* 93, 97.
[43a] The reference is to a series of television programmes with the general title "Beyond Reasonable Doubt", presented during April and May 1977 by Associate Professor Gordon Hawkins of the University of Sydney.
[44] Op. cit. 189, 190.

Despite the material I have presented which suggests that jurors may set a lower standard than the judges when polled individually, the American and English experience reported by Kalven and Zeisel shows clearly that in actual practice juries do require stricter proof than do judges, or at least that for one reason or another they acquit more people than judges would. I have tried to show that the suggestions that have been made for alternative systems of trial based on the use of experts in probabilities are for one reason or another impracticable. There are, of course, other weighty reasons for preferring trial by jury in criminal cases, which I need not elaborate tonight. My concern is with standards of proof, and I cannot bring this lecture to a better conclusion than by quoting from a speech of Lord Devlin, made in Chicago in 1960[45]

> "Trial by jury is not an instrument of getting at the truth; it is a process designed to make it as sure as possible that no innocent man is convicted."

[45] Quoted by Kalven and Zeisel, op. cit. 190.