TRIAL BY MATHEMATICS: PRECISION AND RITUAL IN THE LEGAL PROCESS

Laurence H. Tribe *

Professor Tribe considers the accuracy, appropriateness, and possible dangers of utilizing mathematical methods in the legal process, first in the actual conduct of civil and criminal trials, and then in designing procedures for the trial system as a whole. He concludes that the utility of mathematical methods for these purposes has been greatly exaggerated. Even if mathematical techniques could significantly enhance the accuracy of the trial process, Professor Tribe also shows that their inherent conflict with other important values would be too great to allow their general use.

THE system of legal proof that replaced trial by battle in Continental Europe during the Middle Ages reflected a starkly numerical jurisprudence. The law typically specified how many uncontradicted witnesses were required to establish various categories of propositions, and defined precisely how many witnesses of a particular class or gender were needed to cancel the testimony of a single witness of a more elevated order.[1] So it was that medieval law, nurtured by the abstractions of scholasticism, sought in mathematical precision an escape from the perils of irrational and subjective judgment.

In a more pragmatic era, it should come as no surprise that the search for objectivity in adjudication has taken another tack. Yesterday's practice of numerology has given way to today's theory of probability, currently the *sine qua non* of rational analysis. Without indulging in the dubious speculation that con-

---

* Assistant Professor of Law, Harvard University. A.B. Harvard, 1962; J.D. Harvard, 1966.

[1] *See* M. CAPPELLETTI & J. PERILLO, CIVIL PROCEDURE IN ITALY 35–36 (1965); A. ENGELMANN, A HISTORY OF CONTINENTAL CIVIL PROCEDURE 41–47 (1927); R. GINSBURG & A. BRUZELIUS, CIVIL PROCEDURE IN SWEDEN 33 & n.131, 295 & n.471 (1965); J. GLASER, LEHRE VOM BEWEIS IM STRAFPROZESS 132–35 (1883); Kunert, *Some Observations on the Origin and Structure of Evidence Rules Under the Common Law System and the Civil Law System of "Free Proof" in the German Code of Criminal Procedure*, 16 BUFF. L. REV. 122, 141–42 & nn.99–100, 144–45 (1966). *See also* A. ESMEIN, A HISTORY OF CONTINENTAL CRIMINAL PROCEDURE 264–71 (J. Simpson, transl. 1913); 1 F. HÉLIE, TRAITÉ DE L'INSTRUCTION CRIMINELLE 650–53, 656–57 (1845); F. VOLTAIRE, A COMMENTARY ON BECCARIA'S ESSAY ON CRIMES AND PUNISHMENTS 227–28 (1872).

temporary probabilistic methods will one day seem as quaint as their more mystical predecessors, one can at least observe that the resort to mathematical techniques as alternatives to more intuitive tools in the trial process has ancient roots. Nor is it entirely accidental that those roots seem oddly twisted when examined outside their native soil. For, although the mathematical or pseudo-mathematical devices which a society embraces to rationalize its systems for adjudication may be quite comprehensible to a student of that society's customs and culture, those devices may nonetheless operate to distort — and, in some instances, to destroy — important values which that society means to express or to pursue through the conduct of legal trials. This article discusses the respects in which this is the case — and, in so doing, suggests a framework of analysis for the assessment of the potentialities and dangers of current and proposed uses of mathematical methods in the trial process.

In speaking of mathematical methods "in the trial process," I am referring to two related but nonetheless separable topics: not only to the use of mathematical tools in the actual conduct of a particular trial, but also to the use of such tools in the design of the trial system as a whole. The first topic encompasses such questions as the propriety of allowing the parties in a lawsuit to employ explicitly statistical evidence or overtly probabilistic arguments for various purposes,[2] and the wisdom of permitting or encouraging the trier to resolve the conflicting claims of a lawsuit with the assistance of mathematical methods.[3] The second topic, in contrast, centers on the desirability of employing such methods in establishing the procedural and evidentiary rules according to which lawsuits generally should be conducted. Both topics, of course, share a common core: both involve the wisdom of using mathematical tools to facilitate the making of choices among available courses of action with respect to the trial process. In this sense, both topics form part of the larger subject of when and how mathematical methods ought to

---

[2] I am, of course, aware that *all* factual evidence is ultimately "statistical," and all legal proof ultimately "probabilistic," in the epistemological sense that no conclusion can ever be drawn from empirical data without some step of inductive inference — even if only an inference that things are usually what they are perceived to be. *See, e.g.,* D. HUME, A TREATISE OF HUMAN NATURE, bk. I, pt. III, § 6, at 87 (L.A. Selby-Bigge ed. 1958). My concern, however, is only with types of evidence and modes of proof that bring this probabilistic element of inference to explicit attention in a quantified way. As I hope to show, much turns on whether such explicit quantification is attempted.

[3] By "mathematical methods," I mean the entire family of formal techniques of analysis that build on explicit axiomatic foundations, employ rigorous principles of deduction to construct chains of argument, and rely on symbolic modes of expression calculated to reduce ambiguity to a minimum.

be employed in decisionmaking. And this subject, in turn, is part of the still more inclusive topic of when it is desirable to make decisions in a calculating, deliberate way, with the aid of precise and rigorous techniques of analysis. To the extent that this article sheds any light on those larger matters, I will of course be gratified. I will not, however, attempt to deal directly with them here, and will instead confine myself to the narrower inquiries outlined above.

Two further introductory remarks are in order. First, my subject is the use of mathematics as a tool for decisionmaking rather than simply as a mode of thought, as an instrument rather than as a language. Conceivably, the very enterprise of describing some phenomena in precise mathematical terms, and particularly the enterprise of quantifying them, might be shown to entail some significant costs in addition to its obvious benefits. Perhaps it is in some sense "dehumanizing" to talk in highly abstract or quantitative terms about some subjects,[4] but this is another issue not to be treated here.

Second, although my central concern is the wisdom of using mathematical methods for certain decisionmaking purposes even when those methods are rationally employed, I will also examine what must be regarded as clearly irrational uses of those methods. Thus, some might charge that, by relying on such misuses in any overall assessment, I have confused the avoidable costs of using a tool badly with the inherent costs of using it well. It is rather like the claim that statistics can lie. One may always respond that this claim is false while conceding that the devil can quote Scripture to his own purposes. In a sense, this is obviously the case. But in another sense, it is only a half-truth, for the costs of abusing a technique must be reckoned among the costs of using it at all to the extent that the latter creates risks of the former. To be more precise, in at least some contexts, permitting *any* use of certain mathematical methods entails a sufficiently high risk of misuse, or a risk of misuse sufficiently costly to avoid, that it would be irrational not to take such misuse into account when deciding whether to permit the methods to be employed at all.

Finally, a word about objectives. This analysis has been undertaken partly because I suspect that the lure of objectivity and precision may prove increasingly hard to resist for lawyers concerned with reliable, or simply successful, adjudication; partly because a critique of mathematical efforts to enhance the reliability and impartiality of legal trials may yield helpful insights into what such trials are and ought to be; and partly because such a

---

[4] One senses that much of the contemporary opposition to the technological emphasis upon rationality and technique rests on some such premise.

critique may ultimately contribute to an appreciation of how rigor and quantification, once their real costs and limits are better understood, might actually prove useful in processes of decision-making. Most fundamentally, though, I write in reaction to a growing and bewildering literature of praise for mathematical precision in the trial process,[5] a literature that has tended to catalogue or to assume the virtues of mathematical approaches quite as uncritically as earlier writers [6] tended to deny their relevance.

I. Factfinding with Mathematical Probabilities

A. Mysteries of Expertise

The infamous trial in 1899 of Alfred Dreyfus, Captain in the French General Staff, furnishes one of the earliest reported instances of proof by mathematical probabilities. In attempting to establish that the author of a certain document that allegedly fell into German hands was none other than Captain Dreyfus, the prosecution called several witnesses who theorized that Dreyfus must have written the document in question by tracing the word *intérêt* from a letter written by his brother, constructing a chain of several of these traced words in a row, and then writing over this chain as a model when preparing the document — in order to give it the appearance of a forgery and thereby to protect himself should the document later be traced to him.[7] To identify the writing in the document as that of Dreyfus, the prosecution's witnesses reported a number of close matches between the lengths of certain words and letters in the document and the lengths of certain words and letters in correspondence taken from Dreyfus' home. Obscure lexicographical and graphological "coincidences" within the document itself were said by the witnesses to indicate

---

[5] *See, e.g.,* Cullison, *Probability Analysis of Judicial Fact-Finding: A Preliminary Outline of The Subjective Approach,* 1969 U. Tol. L. Rev. 538 (1969) [hereinafter cited as Cullison]; Finkelstein & Fairley, *A Bayesian Approach to Identification Evidence,* 83 Harv. L. Rev. 489 (1970) [hereinafter cited as Finkelstein & Fairley]. *See also,* Becker, *Crime and Punishment: An Economic Approach,* 76 J. Pol. Econ. 169 (1968) [hereinafter cited as Becker]; Birmingham, *A Model of Criminal Process: Game Theory and Law,* 56 Cornell L. Rev. 57 (1970) [hereinafter cited as Birmingham]; Kaplan, *Decision Theory and the Factfinding Process,* 20 Stan. L. Rev. 1065 (1968) [hereinafter cited as Kaplan]; *cf.* Broun & Kelly, *Playing the Percentages and the Law of Evidence,* 1970 Ill. L. F. 23 [hereinafter cited as Broun & Kelly].

[6] *See, e.g.,* W. Wills, An Essay on the Principles of Circumstantial Evidence 6–10, 15, 282 (4th ed. 1862); M. Houts, From Evidence to Proof 132 (1956).

[7] *See* the trial testimony of Jan. 18, 1899, and Feb. 4, 1899, reported in a special supplement to Le Petit Temps (Paris), April 22, 1899.

the high probability of its disguised character and of its use to convey coded information.[8] To establish the validity of the hypothesis that the document had been traced over the hand-writing of Dreyfus' brother, the prosecution's witnesses computed the "amazing" frequency with which certain letters in the document appeared over the same letters of the word chain constructed by repeating *intérêt* a number of times, once a variety of complex adjustments had been made.[9]

The very opacity of these demonstrations protected them to some degree from effective spontaneous criticism, but the "mathematics" on which they were based was in fact utter nonsense. As the panel of experts appointed several years later to review the evidence in the Dreyfus case easily showed,[10] there was nothing statistically remarkable about the existence of close matches in some word lengths between the disputed document and Dreyfus' correspondence, given the many word pairs from which the prosecution was free to choose those that displayed such similarities.[11] Moreover, the supposed coincidences within the document itself reflected no significant deviation from what one would expect in normal French prose. Finally, the frequency with which various letters in the document could be "localized" over the letters of *intérêt* was likewise statistically insignificant.

Armand Charpentier, a prominent student of the Dreyfus affair, reports that counsel for Dreyfus and the Government Commissioner alike declared that they had understood not a word of the witness' mathematical demonstrations.[12] Charpentier adds that, although the judges who convicted Dreyfus were in all likelihood equally mystified, they nonetheless "allowed themselves to

---

[8] For example, one witness stressed the presence of four coincidences out of the 26 initial and final letters of the 13 repeated polysyllabic words in the document. He evaluated at .2 the probability of an isolated coincidence and calculated a probability of $(0.2)^4 = .0016$ that four such coincidences would occur in normal writing. But $(0.2)^4$ is the probability of four coincidences out of four; that of four or more out of 13 is some 400 times greater, or approximately .7. See *Rappord de Mm. Les Experts Darboux, Appell, et Poincaré*, in LES DOCUMENTS JUDICIARES DE L'AFFAIRE DREYFUS, in LA RÉVISION DU PROCÈS DE RENNES (1909) [hereinafter cited as *Rappord*]. *Cf.* note 40 *infra*.

[9] Two witnesses observed that, when the word chain *"intérêt/intérêt/intérêt/ intérêt . . . ."* was compared with the document itself, allowing one letter of slipping-back for each space between words and aligning the word chain with the actual or the ideal left-hand margin as convenient, the letter *l* appeared with particular frequency over the word-chain letter *i*; the letters *n* and *p* appeared frequently over the word-chain letter *n*; and so on. Far from being in any way remarkable, however, the probability that *some* such pattern can be discerned in any document is nearly certainty. See *Rappord* 534.

[10] *See id.*

[11] *See* the discussion of the "selection effect," note 40 *infra*.

[12] A. CHARPENTIER, THE DREYFUS CASE 52–53 (J. May transl. 1935).

be impressed by the scientific phraseology of the system." [13] It
would be difficult to verify that proposition in the particular case,
but the general point it makes is a crucial one: the very mystery
that surrounds mathematical arguments — the relative obscurity
that makes them at once impenetrable by the layman and impres-
sive to him — creates a continuing risk that he will give such
arguments a credence they may not deserve and a weight they
cannot logically claim.

The California Supreme Court recently perceived this danger
when it warned that "[m]athematics, a veritable sorcerer in our
computerized society, while assisting the trier of fact in the search
for truth, must not [be allowed to] cast a spell over him." [14] The
court ruled improper a prosecutor's misconceived attempt to link
an accused interracial couple with a robbery by using probability
theory. The victim of the robbery, an elderly woman, had testi-
fied that she saw her assailant, a young woman with blond hair,
run from the scene. One of the victim's neighbors had testified
that he saw a Caucasian woman, with her hair in a dark blond
ponytail, run from the scene of the crime and enter a yellow
automobile driven by a male Negro wearing a mustache and
beard. Several days later, officers arrested a couple that seemed
to match these descriptions. [15] At the week-long trial of this
couple, the victim was unable to identify either defendant, and
her neighbor's trial identification of the male defendant was
effectively impeached. [16] Moreover, the defense introduced evi-
dence that the female defendant had worn light-colored clothing
on the day of the robbery, although both witnesses testified that
the girl they observed had worn dark clothing. Finally, both
defendants took the stand to deny any participation in the crime,
providing an alibi that was at least consistent with the testimony
of another defense witness.

In an effort to bolster the identification of the defendants as
the perpetrators of the crime, the prosecutor called a college
mathematics instructor to establish that, if the robbery was indeed
committed by a Caucasian woman with a blond ponytail accom-

---

[13] Id. at 53. See also id. at 265.

[14] People v. Collins, 68 Cal. 2d 319, 320, 438 P.2d 33, 66 Cal. Rptr. 497 (1968).

[15] There was testimony that the female defendant's hair color at the time of
the robbery was light blond rather than dark blond, as it appeared at trial. The
male defendant had no beard at trial or when arrested and told the arresting
offcers that he had not worn one on the day of the robbery. There was testimony
corroborating his claim that he had shaved his beard approximately two weeks
before the robbery, but other testimony that he was bearded the day after the
robbery.

[16] The neighbor admitted at trial "that at the preliminary hearing he [had]
testified to an uncertain identification at the police lineup shortly after the at-
tack . . . ." 68 Cal. 2d at 321, 438 P.2d at 34, 66 Cal. Rptr. at 498.

panied by a Negro with a beard and mustache and driving a yellow car, there was an overwhelming probability that the accused couple were guilty because they matched this detailed description.

The witness first testified to the "product rule" of probability theory, according to which the probability of the joint occurrence of a number of mutually independent events equals the product of the individual probabilities of each of the events.[17] Without presenting any supporting statistical evidence, the prosecutor had the witness assume specific probability factors for each of the six characteristics allegedly shared by the defendants and the guilty couple.[18] Applying the product rule to the assumed factors, the prosecutor concluded that there was but one chance in twelve million that any couple chosen at random would possess the characteristics in question, and asked the jury to infer that there was therefore but one chance in twelve million of the defendants' innocence.

The jury convicted but the California Supreme Court reversed, holding the mathematical testimony and the prosecutor's associated argument inadmissible on four separate grounds. First, the record was devoid of any empirical evidence to support the individual probabilities assumed by the prosecutor.[19]

Second, even if the assumed probabilities were themselves correct, their multiplication under the product rule presupposed the independence of the factors they measured — a presupposition for which no proof was presented, and which was plainly false.[20] If two or more events tend to occur together, the chances of their separate occurrence obviously cannot be multiplied to yield the chance of their joint occurrence.[21] For example, if every tenth

---

[17] See explanation in note 63 infra.

[18]

Characteristic	Assumed Probability of its Occurrence
1. Partly yellow automobile	1/10
2. Man with mustache	1/4
3. Girl with ponytail	1/10
4. Girl with blond hair	1/3
5. Negro man with beard	1/10
6. Interracial couple in car	1/1000

[19] See State v. Sneed, 76 N.M. 349, 414 P.2d 858 (1966); People v. Risley, 214 N.Y. 75, 108 N.E. 200 (1915), discussed at pp. 1344–45 & notes 47–49 infra. See also Campbell v. Board of Educ., 310 F. Supp. 94, 105 (E.D.N.Y. 1970).

[20] The sixth factor, for example, essentially restates parts of the first five. See note 18 supra.

[21] Precisely this mistake is made in C. McCORMICK, HANDBOOK OF THE LAW OF EVIDENCE § 171 (1954) and in J. WIGMORE, THE SCIENCE OF JUDICIAL PROOF § 154, at 270–71 (3d ed. 1937). One court has treated such dependence, I think mistakenly, as going only to the "weight" of the product and not to its admissibility.

man is black and bearded, and if every fourth man wears a mustache, it may nonetheless be true that most bearded black men wear mustaches, so that nearly one man in ten — not one in forty — will be a black man with a beard *and* a mustache.

Third, even if the product rule could properly be applied to conclude that there was but one chance in twelve million that a randomly chosen couple would possess the six features in question, there would remain a substantial possibility that the guilty couple did not in fact possess all of those characteristics — either because the prosecution's witnesses were mistaken or lying, or because the guilty couple was somehow disguised. "Traditionally," the court reasoned, "the jury weighs such risks in evaluating the credibility and probative value of trial testimony," [22] but — finding itself unable to quantify these possibilities of error or falsification — the jury would be forced to exclude such risks from any effort to assign a number to the probability of guilt or innocence and would be tempted to accord disproportionate weight to the prosecution's computations.

Fourth, and entirely apart from the first three objections, the prosecutor erroneously equated the probability that a randomly chosen couple would possess the incriminating characteristics, with the probability that any given couple possessing those characteristics would be innocent. After all, if the suspect population contained, for example, twenty-four million couples, and if there were a probability of one in twelve million that a couple chosen at random from the suspect population would possess the six characteristics in question, then one could well expect to find two such couples in the suspect population, and there would be a probability of approximately one in two — not one in twelve million — that any given couple possessing the six characteristics would be innocent.[23] The court quite reasonably thought that few

---

State v. Coolidge, 109 N.H. 403, 419, 260 A.2d 547, 559 (1969), *cert. granted on other issues*, 399 U.S. 926 (1970) (No. 1318 Misc., 1969 Term; renumbered No. 323, 1970 Term), *discussed at note 40 infra*.

[22] 68 Cal. 2d at 330, 438 P.2d at 40, 66 Cal. Rptr. at 504.

[23] In a separate mathematical appendix, the court demonstrated that, even if the number of suspect couples approaches only twelve million, the probability that *at least one other couple* (in addition to the actually guilty couple) will possess the six characteristics rises to somewhat over *forty-one percent*, even on the assumption that the prosecutor was correct in concluding that the probability that a randomly chosen couple would possess all six characteristics is but one in twelve million. More generally, the court showed that the probability of such duplication equals

$$\frac{1-(1-Pr)^N - NPr(1-Pr)^{N-1}}{1-(1-Pr)^N},$$

where Pr equals the probability that a random couple will possess the characteristics in question and N is the number of couples in the suspect population. 68 Cal. 2d at 333–35, 438 P.2d at 42–43, 66 Cal. Rptr. at 506–07. If λ is taken to represent the value of $N \cdot Pr$, then the Poisson approximation for the above

defense attorneys, and fewer jurors, could be expected to comprehend these basic flaws in the prosecution's analysis.[24] Under the circumstances, the court concluded, this "trial by mathematics" so distorted the jury's role and so disadvantaged defense counsel as to constitute a miscarriage of justice.[25]

But the California Supreme Court discerned "no inherent incompatability between the disciplines of law and mathematics and intend[ed] no general disapproval . . . of the latter as an auxiliary in the fact-finding processes of the former." [26]   Thus expressed, the court's position seems reasonable enough. Any highly specialized category of knowledge or technique of analysis is likely to share in some degree the divergence between impressiveness and understandability that characterizes mathematical proof; surely, adjudication should not for that reason be deprived

---

quotient is $1 - \dfrac{\lambda}{e^\lambda - 1}$, where e is the transcendental number 2.71828. . . . that is used as the base for natural logarithms. See 1 W. FELLER, AN INTRODUCTION TO PROBABILITY THEORY AND ITS APPLICATIONS 153–64 (3d ed. 1968); Kingston, *Applications of Probability Theory in Criminalistics*, 60 J. AM. STATIST. ASS'N 70, 74 (1965). On the assumption that $Pr = 1/N$ (so that $\lambda = 1$), the value of the above quotient as N grows without limit is thus $(e-2)/(e-1)$, which is approximately .42, as the court correctly concluded. See Cullison, *Identification by Probabilities and Trial by Arithmetic (A Lesson For Beginners in How to be Wrong With Greater Precision)*, 6 HOUST. L. REV. 471, 484–502 (1969).

Finkelstein and Fairley suggest that the court's argument was mathematically incorrect because "the court's assumption that one in twelve million is a fair estimate of the probability of selecting such a couple at random necessarily implies that it is a fair estimate of the number of such couples in the population." Finkelstein & Fairley 493; *accord*, Broun & Kelly, *supra* note 5, at 43. But this completely misconceives the argument. Of course, if the figure of one in twelve million had represented an estimate, based upon random sampling, of *the actual frequency of Collins-like couples in a known population*, the criticism of the court's opinion would be well taken. But in fact the "one-in-twelve-million" figure represented nothing of the sort. Since nothing was known about exactly who was and who was not a member of the population of "suspect" couples, that figure represented only an estimate of the probability that any given couple, chosen at random from an unknown population of "suspect" couples, would turn out to have the six "Collins" characteristics — with that estimate itself based only on a multiplication of component factors, each representing the frequency of one of the six characteristics in a much larger population.

[24] *See also* note 40 *infra*.

[25] The court stressed the fact that the prosecutor had criticized the traditional notion of proof beyond a reasonable doubt as "hackneyed" and "trite"; that he "sought to reconcile the jury to the risk that, under his 'new math' approach to criminal jurisprudence, 'on some rare occasion . . . an innocent person may be convicted' "; and that he thereby sought "to persuade the jury to convict [the] defendants whether or not they were convinced of their guilt to a moral certainty and beyond a reasonable doubt." 68 Cal. 2d at 331–32, 438 P.2d at 41, 66 Cal. Rptr. at 505. The interaction between mathematical proof and reasonable doubt is discussed at pp. 1372–75 *infra*.

[26] 68 Cal. 2d at 320, 438 P.2d at 33, 66 Cal. Rptr. at 497.

of the benefits of all expertise. On the contrary, the drawing of unwarranted inferences from expert testimony has long been viewed as rectifiable by cross-examination, coupled with the opportunity to rebut. Particularly if these devices are linked to judicial power to give cautionary jury instructions and to exclude evidence altogether on a case-by-case basis if prejudicial impact is found to outweigh probative force, and if these techniques are then supplemented by a requirement of advance notice of intent to use a particular item of technical proof, and by some provision for publicly financed expert assistance to the indigent accused confronted with an expert adversary,[27] there might seem to be no valid remaining objection to probabilistic proof.

But can such proof simply be equated with expert evidence generally, or does it in fact pose problems of a more pervasive and fundamental character? A consideration of that question requires the more careful development of just what "mathematical proof" should be taken to mean, and what major forms it can assume.

B. Illustrative Cases: Occurrence; Identity; Intention

In an examination of the role of mathematical methods in the trial itself, whether used by one or more of the parties in the presentation of proof or employed by the trier in reaching a decision, we may set aside at the outset those situations in which the very issues at stake in a litigation are in some sense mathematical and hence require the explicit trial use of mathematical techniques — when, for example, the governing substantive law makes a controversy turn on such questions as percentage of market control,[28] expected lifetime earnings,[29] likelihood of widespread public confusion,[30] or the randomness of a jury selection process.[31] My concern is with cases in which mathematical methods are turned to the task of deciding what occurred on a particular, unique occasion, as opposed to cases in which the very

---

[27] See, e.g., 1967 DUKE L.J. 665, 681–83, discussing State v. Sneed, 76 N.M. 349, 414 P.2d 858 (1966).

[28] E.g., United States v. United Shoe Machinery Corp., 110 F. Supp. 295, 304–05 (D. Mass. 1953) (Wyzanski, J.), aff'd per curiam, 347 U.S. 521 (1954).

[29] See, e.g., Louisville & N.R.R. v. Steel, 257 Ala. 474, 59 So. 2d 664 (1952); Von Tersch v. Ahrendsen, 251 Iowa 115, 99 N.W.2d 287 (1959).

[30] See, e.g., United States v. 88 Cases, More or Less, Containing Bireley's Orange Beverage, 187 F.2d 967, 974 (3d Cir. 1951).

[31] See generally, Finkelstein, The Application of Statistical Decision Theory to the Jury Discrimination Cases, 80 HARV. L. REV. 338 (1966); Zeisel, Dr. Spock and the Case of the Vanishing Women Jurors, 37 U. CHI. L. REV. 1 (1969) [hereinafter cited as Zeisel]. But see State v. Smith, 102 N.J. Super. 325, 341, 246 A.2d 35, 50 (1968), aff'd, 55 N.J. 476, 262 A.2d 868 (1970), cert. denied, 400 U.S. 949 (1970).

task defined by the applicable law is that of measuring the sta-
tistical characteristics or likely effects of some process or the
statistical features of some population of people or events.

With this initial qualification in mind, it is possible — and
will occasionally prove helpful — to separate mathematical proof
into three distinct but partially overlapping categories: (1) those
in which such proof is directed to the *occurrence* or nonoccurrence
of the event, act, or type of conduct on which the litigation is
premised; (2) those in which such proof is directed to the *iden-
tity* of the individual responsible for a certain act or set of acts;
and (3) those in which such proof is directed to *intention* or to
some other mental element of responsibility, such as knowledge
or provocation. In dealing with the utility of mathematical proof
in the trial process, I will later show how such a tripartite divi-
sion can be useful. It is sufficient to say at this stage that the
significance, appropriateness, and dangers of mathematical proof
may depend dramatically on whether such proof is meant to bear
upon occurrence, identity, or frame of mind.[32] Several examples
should suffice to illustrate the contents of each of these categories.

1. Occurrence. — Consider first the cases in which the exist-
ence of the legally significant occurrence or act is itself in
question. A barrel falls from the defendant's window onto the
plaintiff's head. The question is whether some negligent act or
omission by defendant caused the fall. Proof is available to
support a finding that, in over sixty percent of all such barrel-
falling incidents, a negligent act or omission was the cause.
Should such proof be allowed and, if so, to what effect? [33]

A man is found in possession of heroin. The question is
whether he is guilty of concealing an illegally imported narcotic
drug. Evidence exists to support the finding that ninety-eight

---

[32] *See* pp. 1365–67, p. 1381 & notes 33, 37 & 41 *infra.*

[33] A sensible, and now quite conventional, approach to this question is "to
treat the probability as the fact if the defendant has the power to rebut the in-
ference." Jaffe, *Res Ipsa Loquitur Vindicated,* 1 BUFF. L. REV. 1, 6 (1951). On
this theory, if the defendant produces a reasonably satisfactory explanation con-
sistent with a conclusion of no negligence, and if the plaintiff produces no further
evidence, the plaintiff should lose on a directed verdict despite his mathematical
proof — unless (1) he can adequately explain his inability to make a more par-
ticularized showing (a possibility not adverted to in *id.*), or (2) no specific explana-
tion is given, but there is some policy reason to ground liability in the area in ques-
tion on a substantial probability of negligence in the *type* of case rather than to re-
quire a reasoned probability in the *particular* case, *cf.* note 100 *infra,* thereby moving
toward a broader basis of liability. It will be noticed that no such policy is likely to
operate when the mathematical evidence goes to the question of the defendant's
identity and the plaintiff does not explain his failure to produce any more par-
ticularized evidence, for it will almost always be important to impose liability
on the correct party, *whatever* the basis of such liability might be. *See* p. 1349
infra. See also notes 37 & 102 *infra.*

percent of all heroin in the United States is illegally imported. What role, if any, may that fact play at the defendant's trial? [34]

A man is charged with overtime parking in a one-hour zone. The question is whether his car had remained in the parking space beyond the time limit. To prove that it had not been moved, the government calls an officer to testify that he recorded the positions of the tire air-valves on one side of the car. Both before and after a period in excess of one hour, the front-wheel valve was pointing at one o'clock; the rear-wheel valve, at eight o'clock. The driver's defense is that he had driven away during the period in question and just happened to return to the same parking place with his tires in approximately the same position. The probability of such a fortunate accident is somewhere between one in twelve and one in one hundred forty-four.[35] Should proof of that fact be allowed and, if so, to what end? [36]

2. *Identity.* — Consider next the cases in which the identity of the responsible agent is in doubt. Plaintiff is negligently run down by a blue bus. The question is whether the bus belonged to the defendant. Plaintiff is prepared to prove that defendant

---

[34] It has now been settled as a federal constitutional matter, *see* Turner v. United States, 396 U.S. 398 (1970), that this statistical fact permits a legislature to authorize a jury to find illegal importation once it finds possession "unless the defendant explains the possession to the satisfaction of the jury." 21 U.S.C. § 174 (1964); *cf.* Leary v. United States, 395 U.S. 6 (1969). At least one commentator has urged the alternative position that the jury should not in such cases be in-structed that proof of possession is sufficient to find illegal importation (for that shifts to the accused the practical burden of persuasion, with its accompanying pressure to testify, notwithstanding any contrary jury charge) but should instead be told that ninety-eight percent of all heroin in the United States is illegally im-ported (for that leaves the jury more likely to give even the non-testifying accused the benefit of the doubt created by the remaining two percent). Comment, *Statu-tory Criminal Presumptions: Reconciling the Practical With the Sacrosanct,* 18 U.C.L.A. L. REV. 157 (1970). But it is by no means clear, despite the commenta-tor's assertion, that "the jury is more likely to consider other relevant circumstances unique to the particular case on a more equal footing with the 98 percent statistic than it would with a presumption." *Id.* at 183 n.102. *See generally* the discussion at pp. 1359-65 *infra.*

[35] If tires rotated in complete synchrony with one another, the probability would be 1/12; if independently, 1/12 × 1/12, or 1/144.

[36] A Swedish court, computing the probability at 1/12 × 1/12 = 1/144 on the dubious assumption that car wheels rotate independently, ruled that fraction large enough to establish reasonable doubt. Parkeringsfrägor, II. Tilförlitligheten av det S.K. locksystemet för parkernigskontroll. *Svensk juristidining,* 47 (1962) 17–32, *cited in* Zeisel, *supra* note 31, at 12. The court's mathematical knife cut both ways, however, for it added that, had all four tire-valves been recorded and found in the same position, the probability of 1/12 × 1/12 × 1/12 × 1/12 = 1/20, 736 would have constituted proof beyond a reasonable doubt. *Id.* For a discussion of why no such translation of the "reasonable doubt" concept into mathematical terms should be attempted, see pp. 1372–75 *infra.*

operates four-fifths of all the blue buses in town. What effect, if any, should such proof be given? [37]

A policeman is seen assaulting someone at an undetermined time between 7 p.m. and midnight. The question is whether the defendant, whose beat includes the place of the assault, was the particular policeman who committed the crime. It can be shown that the defendant's beat brings him to the place of the assault four times during the relevant five-hour period each night, and that other policemen are there only once during the same period. In what way, if at all, may this evidence be used? [38]

A man is found shot to death in the apartment occupied by his mistress. The question is whether she shot him. Evidence is available to the effect that, in ninety-five percent of all known cases in which a man was killed in his mistress' apartment, the mistress was the killer. How, if at all, may such evidence be used? [39]

A civil rights worker is beaten savagely by a completely bald

---

[37] In Smith v. Rapid Transit, Inc., 317 Mass. 469, 58 N.E.2d 754 (1945), the actual case on which this famous chestnut is based, no statistical data were in fact presented, but the plaintiff did introduce evidence sufficient to show that the defendant's bus line was the only one chartered to operate on the street where the accident occurred. Affirming the direction of a verdict for the defendant, the court observed: "The most that can be said of the evidence in the instant case is that perhaps the mathematical chances somewhat favor the proposition that a bus of the defendant caused the accident. This was not enough." 317 Mass. at 470, 58 N.E.2d at 755. *See also* Sawyer v. United States, 148 F. Supp. 877 (M.D. Ga. 1956); Reid v. San Pedro, L.A.&S.L.R.R., 39 Utah 617, 118 P. 1009 (1911). If understood as insisting on a numerically higher showing — an "extra margin" of probability above, say, .55 — then the decision in *Smith* would make no sense, at least if the court's objective were the minimization of the total number of judicial errors in situations of this kind, an objective essentially implicit in the adoption of a "preponderance of the evidence" standard. *See* Ball, *The Moment of Truth: Probability Theory and Standards of Proof,* 14 Vand. L. Rev. 807, 822–23 (1961) [hereinafter cited as Ball]. But cases like *Smith* are entirely sensible if understood instead as insisting on the presentation of *some* non-statistical and "individualized" proof of identity before compelling a party to pay damages, and even before compelling him to come forward with defensive evidence, absent an adequate explanation of the failure to present such individualized proof. *Compare* p. 1349 *infra with* note 33 *supra.*

[38] Note that in this criminal case, as in the preceding civil one, *a fact known about the particular defendant* provides reason to believe that the defendant is involved in a certain percentage of all cases (here, cases of being at the crucial place between 7 p.m. and midnight) possessing a characteristic shared by the litigated case.

[39] In this case, unlike the preceding two, it is *a fact known about the particular event that underlies the litigation,* not any fact known about the defendant, that triggers the probabilistic showing: a certain percentage of all events in which the crucial fact (here, the killing of a man in his mistress' apartment) is true are supposedly caused by a person with a characteristic (here, being the mistress) shared by the defendant in this case.

man with a wooden left leg, wearing a black patch over his right eye and bearing a six-inch scar under his left, who flees from the scene of the crime in a chartreuse Thunderbird with two dented fenders. A man having these six characteristics is charged with criminal battery. The question is whether the defendant is in fact the assailant. Evidence is available to show that less than one person in twenty has any of these six characteristics, and that the six are statistically independent, so that less than one person in sixty-four million shares all six of them. In what ways, if at all, may that calculation be employed? [40]

3. Intention. — Consider finally the cases in which the issue is one of intent, knowledge, or some other "mental" element of responsibility. A recently insured building burns down. The insured admits causing the fire but insists that it was an accident.

---

[40] This is, of course, People v. Collins, 68 Cal. 2d 319, 438 P.2d 33, 66 Cal. Rptr. 497 (1968), minus the specific mathematical errors of *Collins* and without the interracial couple. One special factor that can lead to major mathematical distortions in this type of case is the "selection effect" that may arise from either party's power to choose matching features for quantification while ignoring non-matching features, thereby producing a grossly exaggerated estimate of the improbability that the observed matching would have occurred by chance. *See* Finkelstein & Fairley 495 n.14. This difficulty may well have been present in People v. Trujillo, 32 Cal. 2d 105, 194 P.2d 681, *cert. denied*, 335 U.S. 887 (1948), in which an expert examined a large number of fibers taken from clothing worn by the accused and concluded, upon finding eleven matches with fibers taken from the scene of the crime, that there was only a one-in-a-billion probability of such matching occurring by chance. A particularly egregious case of this sort is State v. Coolidge, 109 N.H. 403, 260 A.2d 547 (1969), *cert. granted on other issues*, 399 U.S. 926 (1970) (No. 1318 Misc., 1969 Term; renumbered No. 323, 1970 Term), where particles taken from the victim's clothing were found to match particles taken from the defendant's car and clothing in twenty-seven out of forty cases.

> In expressing his conclusion based upon statistical probabilities, the [consultant in micro-analysis and director of a university laboratory for scientific investigation] relied upon previous studies made by him, indicating that the probability of finding similar particles in sweepings from a series of automobiles was one in ten. Applying this as a standard, he determined the probability of finding 27 similar particles in sweepings from independent sources would be only one in ten to the 27th power.

109 N.H. at 418–19, 260 A.2d at 559. The court upheld the admissibility of that testimony, 109 N.H. at 422, 260 A.2d at 561, notwithstanding the weakness of the underlying figure of 1/10 and the expert's own concession that the particle sweepings "may not have been wholly independent," 109 N.H. at 419, 260 A.2d at 559. *See* note 21 *supra*. Most significantly, the court was evidently unaware that the relevant probability, that of finding 27 or more matches *out of 40 attempts,* was very much larger than $1/10^{27}$ — larger, in fact, by a factor of approximately 10^{10}. Indeed, even the 40 particles chosen for comparison were visually selected for similarity from a still larger set of particle candidates, 109 N.H. at 421, 260 A.2d at 560 — so large a set, conceivably, that the probability of finding 27 or more matches in sweeping over such a large sample, even from two entirely different sources, could well have been as high as 1/2 or more. *Cf.* note 8 *supra*. Oddly, the expert testimony in *Coolidge* has recently been described as "not misleading." Broun & Kelly, *supra* note 5, at 48.

On the question of intent to commit arson, what use, if any, may be made of evidence tending to show that less than one such fire out of twenty is in fact accidentally caused? [41]

As in an earlier example,[42] a man is found possessing heroin. This time the heroin is stipulated at trial to have been illegally imported. In his prosecution for concealing the heroin with knowledge that it had been illegally imported, what effect may be given to proof that ninety-eight percent of all heroin in the United States is in fact illegally imported? [43]

A doctor sued for malpractice is accused of having dispensed a drug without adequate warning, knowing of its tendency to cause blindness in pregnant women. Should he be allowed to introduce evidence that ninety-eight percent of all doctors are unaware of the side-effect in question?

4. An Overview. — The reader will surely note that this collection of cases might have been subdivided along a variety of different axes. Some of the cases are civil, others criminal. Some involve imputations of moral fault; others do not. Some rest upon statistical calculations that might readily be made; others, on figures that are at best difficult to obtain and at worst entirely inaccessible. Some entail the use of probabilistic evidence to establish liability; others, to negate it. In some, the probabilities refer to a party's own involvement in a category of events; in others, they refer to the proportion of similar events in which a certain critical feature is present, or in which the responsible party has a certain important characteristic. In some of these cases, the mathematics seems best suited to assisting the judge in his allocation of burdens of production or persuasion; in others, its most natural role seems to be as evidence for the finder of fact.

My aim in classifying the cases in terms of occurrence, identity, and intention is not to imply that these other ways of carving up the topic have less significance, but merely to sketch one possible map of the territory I mean to cover — using a set of boundaries that are intuitively suggestive and that will prove helpful from time to time as the discussion unfolds.[44]

Courts confronted with problems of the several sorts enumerated in the three preceding sub-sections have reacted to them on an almost totally ad hoc basis, occasionally upholding an attempt

---

[41] It is, of course, a fair question how such evidence could ever be compiled; the difficulty, and perhaps the impossibility, of compiling it no doubt reflects the "non-objective" nature of the intent inquiry. *See* pp. 1365–66 *infra*.

[42] *See* p. 1339 *supra*.

[43] Turner v. United States, 396 U.S. 398 (1970), sustained an authorized jury inference of knowledge in these circumstances. *See* note 34 *supra*.

[44] *See* note 32 *supra*.

at probabilistic proof,[45] but more commonly ruling the particular attempt improper.[46] A perhaps understandable pre-occupation with the novelties and factual nuances of the particular cases has marked the opinions in this field, to the virtual exclusion of any broader analysis of what mathematics can or cannot achieve at trial — and at what price. As the number and variety of cases continue to mount, the difficulty of dealing intelligently with them in the absence of any coherent theory is becoming increasingly apparent. Believing that a more general analysis than the cases provide is therefore called for, I begin by examining — and ultimately rejecting — the several arguments most commonly advanced against mathematical proof. I then undertake an assessment of what I regard as the real costs of such proof, and reach several tentative conclusions about the balance of costs and benefits.

C. The Traditional Objections

The cases sketched in the preceding section differ in many respects, but all of them share three central features which have at times been thought to preclude any meaningful application of mathematical techniques. The first of those is that, in all of these cases, concepts of probability based upon the relative frequency of various events must be applied, if at all, not to the statistical prediction of a possible *future* event but to a determination of the occurrence or characteristics of an alleged *past* event. At first glance, probability concepts might appear to have no application in deciding precisely what did or did not happen on a specific prior occasion: either it did or it didn't — period.

The New York Court of Appeals elevated that intuition into a rule of law when it rejected probabilistic testimony to show that a forgery had been done on the defendant's typewriter.[47]

---

[45] *See, e.g.*, People v. Trujillo, 32 Cal. 2d 105, 194 P.2d 681, *cert. denied*, 335 U.S. 887 (1948); State v. Coolidge, 109 N.H. 403, 260 A.2d 547 (1969), *cert. granted on other issues*, 399 U.S. 926 (1970) (No. 1318 Misc., 1969 Term; Renumbered No. 323, 1970 Term), *discussed in* note 40 *supra. See also* Note, *The Howland Will Case*, 4 AM. L. REV. 625 (1870), *discussing* Robinson v. Mandell, 20 F. Cas. 1027 (No. 11,959) (C.C.D. Mass. 1868), *discussed in* note 47 *infra*; People v. Jordan, 45 Cal. 2d 697, 707, 290 P.2d 484, 490 (1955), *discussed in* note 155 *infra*.

[46] *See, e.g.*, People v. Collins, 68 Cal. 2d 319, 438 P.2d 33, 66 Cal. Rptr. 497 (1968), *discussed at* pp. 1334–37 *supra*; State v. Sneed, 76 N.M. 349, 414 P.2d 858 (1966); People v. Risley, 214 N.Y. 75, 108 N.E. 200 (1915). *See also* Smith v. Rapid Transit, Inc., 317 Mass. 469, 58 N.E.2d 754 (1945), *discussed in* note 37 *supra*; Miller v. State, 240 Ark. 340, 399 S.W.2d 268 (1966), *discussed in* note 155 *infra*.

[47] People v. Risley, 214 N.Y. 75, 108 N.E. 200 (1915). Experts on typewriters had been called to testify that certain peculiarities in the forged document corre-

The court distinguished the judicially accepted use of life ex-
pectancy tables on the ground that such use arises "from neces-
sity when the fact to be proved is the probability of the happening
of a future event. It would not be allowed," the court continued,
"if the fact to be established were whether A had in fact died,
to prove by the Carlisle Table he should still be alive." [48] Thus,
the court reasoned, probabilistic testimony (as to the rarity of
the coincidence between peculiarities in the defendant's typewriter
and peculiarities in the forged document) should be disallowed
since the "fact to be established . . . was not the probability of
a future event, but whether an occurrence asserted by the people
to have happened had actually taken place." [49] The court's re-
sult was defensible on far narrower grounds,[50] but this reasoning
is not. It is not the future character of an event that induces us
to give weight to probabilistic evidence, but the lack of other,
more convincing, evidence — an absence more common in, but

---

sponded completely with peculiarities in a typed sample produced by the defendant's
typewriter. Cf. State v. Freshwater, 30 Utah 442, 85 P. 447 (1906). A mathe-
matician was then allowed to testify, in response to a hypothetical question ascrib-
ing certain probabilities to the occurrence of any one defect in a random typewriter,
that the probability of the coincidence of all these defects in any single machine
was but one in four billion. Cf. Note, supra note 45, at 648-49 (1870), discussing
Robinson v. Mandell, 20 F. Cas. 1027 (No. 11,959) (C.C.D. Mass 1868), in which
Benjamin Pierce, Harvard Professor of Mathematics, applied the product rule to
strokes of authentic and disputed signatures to conclude that their similarities should
be expected to occur by chance only once in a number of times equal to the
thirtieth power of five. The Risley court found reversible error in the use of the
"one-in-four-billion" argument on the narrow and obviously correct ground that
the hypothetically assumed probabilities for the separate defects were unsupported
by any evidence in the case. But the court went on to indicate its view, quoted
in text above, that probabilistic evidence is necessarily inadmissible to establish a
past event.
    [48] 214 N.Y. at 86, 108 N.E. at 203. If the court's use of the phrase "should
still be alive" is taken to suggest that A's death has otherwise been firmly estab-
lished, then the court's example has a surface plausibility arising not out of the fact
that A's alleged death is past rather than future, but out of the obviously low proba-
tive force of general statistical averages of this sort when confronted with convinc-
ing evidence more narrowly focused on the disputed event itself. Cf. note 100
infra. If, on the other hand, the court's assertion is taken to deny the relevance
of life expectancy data in deciding a genuine factual dispute as to whether or not
A had died, then the court's denial flies in the face of at least the theory that
underlies the traditional presumption of death in cases of long and unexplained
absence. See Comment, A Review of the Presumption of Death in New York, 26
ALBANY L. REV. 231, 245 (1962). See also note 101 infra.
    [49] 214 N.Y. at 86, 108 N.E. at 203. But see Liddle, Mathematical and Statistical
Probability As a Test of Circumstantial Evidence, 19 CASE W. RES. L. REV. 254,
277-78 (1968), expressing the surprising view that "[m]athematical probability
is . . . most useful in establishing the existence of or identifying facts relating to
past events and least useful in the predicting of future events . . . ."
    [50] See note 47 supra.

certainly not limited to, future occurrences.[51] Indeed, "sense-perception itself" might be viewed as "a form of prediction for action purposes," [52] and "propositions about past facts . . . [can be regarded as] 'predictions,' on existing information, as to what the 'truth' will turn out to be when and if more knowledge is available." [53] Insofar as the relevance of probability concepts is concerned, then, there is simply no inherent distinction between future and past events. That all of the cases sketched in the preceding section would apply such concepts to a determination about a past occurrence therefore gives rise to no objection of substance.

However, a second similarity among the cases put above is less easily dismissed: in all of them, making use of the mathematical information available first requires transforming it from evidence about the *generality* of cases to evidence about the *particular* case before us. Some might suggest that no such transformation is possible, and that no translation can be made from probability as a measure of objective frequency in the generality of cases to probability as a measure of subjective belief in the particular instance. That suggestion would be incorrect.[54] In the bus case, to take a typical example, we start with the objective fact that four-fifths of the blue buses are operated by the defendant. That datum can obviously point to a correct conclusion in the particular case, for it suggests that, in the absence of other information, in some sense there is a "four-fifths certainty" that the defendant's bus hit this plaintiff. To be sure, the complete "absence of other information" is rare,[55] but the mathematical datum nonetheless provides a useful sort of knowledge — in part to guide the judge's allocation of the burden of producing believable evidence,[56] and in part to convey to the factfinder a relatively precise sense of the probative force of the background information that is available.

But does it really mean anything at all to be "four-fifths

---

[51] *See* Note, *Evidential Use of Mathematically Determined Probability*, 28 HARV. L. REV. 693, 695–96 (1915).

[52] Ball, *supra* note 37, at 815 n.19, *citing* A. AMES, THE MORNING NOTES OF ADELBERT AMES, JR. (H. Cantril ed. 1960).

[53] Ball, *supra* note 37, at 815.

[54] *Cf.* note 100 *infra* for a related but more supportable proposition.

[55] If no other evidence has been adduced by the time the plaintiff has rested, we at least know the very fact that no other evidence has been adduced — a fact that may properly be treated as dispositive in some situations if other evidence on the issue of identity seems likely to have been available. *Cf.* Case v. New York Central R.R., 329 F.2d 936, 938 (2d Cir. 1964) (Friendly, J.); National Life & Accident Ins. Co. v. Eddings, 188 Tenn. 512, 221 S.W.2d 695 (1949). *See* p. 1349 *infra*.

[56] *See* note 33 *supra* and note 102 *infra*.

1971] TRIAL BY MATHEMATICS certain" in a particular case? Unlike many such questions, this last, fortunately, has an answer — one first formulated rigorously by Leonard Savage in a seminal 1950 work.[57] Professor Savage, employing elegantly few assumptions, developed a "personalistic" or "subjective" theory of probability based on the notion that it makes sense to ask someone what he would do if offered a reward for guessing correctly whether any proposition, designated X, is true or false. If he guesses that X is true under these circumstances, we say that *for him* the subjective probability of X, written $P(X)$, exceeds fifty percent. Symbolically, $P(X) > .5$. If he would be equally satisfied guessing either way, then we say that, for him, $P(X) = .5$.

As Professor Savage demonstrated, this basic concept can readily be expanded into a complete scale of probabilities ranging from 0 to 1, with $P(X) = 0$ representing a subjective belief that X is impossible and $P(X) = 1$ representing a subjective belief that X is certain.[58] Thus, one could take a sequence of boxes each containing a well-shuffled deck of one hundred cards, some marked "True" and others marked "False." The first box, B_0, would contain no cards marked "True" and 100 marked "False"; the second box, B_1, would contain 1 card marked "True" and 99 marked "False;" the next box, B_2, would contain 2 cards marked "True" and 98 marked "False"; and so on, until the last box, B_{100}, would contain 100 cards marked "True" and no cards marked "False." To say that a person is "four-fifths" or "eighty percent" certain of X is simply to say that he would be as willing to bet that the proposition X is true as he would be willing to bet (with the same stakes) that a card chosen at random from box B_{80} will turn out to be marked "True." In these circumstances, the person would say that, for him, $P(X) = .8$.

In the context of the bus case, if a person knew only that eighty percent of all blue buses are operated by the defendant,[59] and if he had to bet one way or the other, he should be as willing to bet that the bus involved *in this case* was defendant's as he would be willing to bet that a card chosen at random from B_{80}

---

[57] L. SAVAGE, FOUNDATIONS OF STATISTICS (1950).

[58] The notion of probabilities as measures of an individual's "degree of confidence" or "degree of belief" in an uncertain proposition or event traces to JAMES BERNOULLI, ARS CONJECTANDI (1713). See the extremely helpful historical discussion in H. RAIFFA, DECISION ANALYSIS, INTRODUCTORY LECTURES ON CHOICES UNDER UNCERTAINTY 273–78 (1968) [hereinafter cited as RAIFFA]. The special contribution of Savage was to formalize that notion in terms useful for the rigorous study of decisionmaking under uncertainty.

[59] It is possible to know *only* this fact at the outset of the trial — though not, of course, at its conclusion, *see* note 55 *supra* — unless the bulk of the trial record were somehow destroyed and, with it, all memory of what had been established during the proceedings.

This content downloaded from 128.237.186.189 on Thu, 29 Dec 2016 18:22:41 UTC
All use subject to http://about.jstor.org/terms

would be marked "True." [60]   This is what it means to say that, for him, the subjective probability of the defendant's liability [61] — pending the receipt of further information — equals .8.

Different people, of course, would typically assign different subjective probabilities to the same propositions — but that is as it must be, unless the propositions in question are unusually simple. And at least in the law we do not find startling the no- tion that reasonable men with differing life experiences and differ- ing assumptions will assess the same evidence differently.

The interesting thing about subjective probabilities as defined by Savage is that, once a few entirely plausible postulates are accepted, [62] these probabilities obey the usual rules that the school- boy associates with such simple operations as flipping fair coins or drawing cards from a well-shuffled deck; hence the translation from objective frequencies to subjective probabilities called for by all of the cases we have considered can indeed be made. [63]

---

[60] To make the idea of a "bet" involving B_{80} correspond as nearly as possible to the situation that actually confronts the trier of fact, one need only postulate that the reward accorded a correct guess consists in learning that a particular lawsuit which the trier wants to see rightly decided has in fact been correctly deter- mined.

[61] Equating the defendant's liability with the mere fact of identity may, of course, overlook other important elements of legal responsibility, a matter taken up at pp. 1365–66 *infra*. I make the equation here only on the explicit assump- tion that identity is the sole issue in the litigation.

[62] Typical is the "transitivity" postulate that one who regards X as more probable than Y, and Y as more probable than Z, should also regard X as more probable than Z. The other postulates simply specify that $P(X)$ is never less than zero or greater than one and that, if A and B are any two mutually exclusive propositions, $P(A) + P(B)$ equals $P(A$ or $B)$.

[63] Perhaps the most involved of the few basic rules that make the trans- lation a useful one is the rule used to calculate the probability that two proposi- tions, A and B, are both true. If a person learned somehow that A was in fact true, how sure would he *then* be of B's truth? The answer to that question, cali- brated in terms of the sequence of boxes described above, measures the probability, for this person, of B *conditioned on A*, or of B *given A*, written $P(B|A)$. See note 71 *infra*. The rule in question simply states that the person's probability estimate of the *joint* truth of A and B equals his probability estimate of A, multiplied by his probability estimate of B conditioned on A. Symbolically, $P(A\&B) = P(A) \cdot P(B|A)$. For example, if a person thinks that A is exactly as likely to be true as false (*i.e.*, $P(A) = 1/2$), and if learning of A's truth would lead him to think that B is only half as likely to be true as false (*i.e.*, $P(B|A) = 1/3$), then he should con- clude that the probability of A and B *both* being true is $(1/2) \cdot (1/3) = 1/6$. In the special case where he believes A and B are *mutually independent*, his knowl- edge of A's truth would have no effect on his estimate of B, so that, for him, $P(B|A)$ would equal $P(B)$, and thus $P(A\&B)$ would simply equal $P(A) \cdot P(B)$ — which is the "product rule," *noted at* p. 1335 *supra*.

It should be noted here that Finkelstein and Fairley suggest that no complete translation from objective frequencies to subjective probabilities is actually re- quired, for they theorize that even subjective probabilities may be interpreted as

Again, there are no insuperable obstacles to the application of mathematical techniques.

Once the translation to subjective probabilities is completed, we encounter the third similarity among the cases. In very few of them, if any, can the mathematical evidence, *taken alone and in the setting of a completed lawsuit*, establish the proposition to which it is directed with sufficient probative force to prevail. To return once again to the blue bus litigation,[64] even assuming a standard of proof under which the plaintiff need only establish his case "by a preponderance of the evidence" in order to suc- ceed, the plaintiff does not discharge that burden by showing simply that four-fifths, or indeed ninety-nine percent, of all blue buses belong to the defendant.[65] For, unless there is a satisfactory explanation for the plaintiff's singular failure to do more than present this sort of general statistical evidence, we might well rationally arrive, once the trial is over, at *a subjective probability of less than .5* that defendant's bus was really involved in the specific case.[66] And in any event, absent satisfactory explana- tion, there are compelling reasons of policy to *treat* the subjective probability as less than .5 — or simply as insufficient to support a verdict for plaintiff. To give less force to the plaintiff's evi- dentiary omission would eliminate any incentive for plaintiffs to do more than establish the background statistics. The upshot would be a regime in which the company owning four-fifths of

---

expressing frequencies. Thus, they equate the statement that the subjective probability of the defendant's guilt is 1/2 with the statement that "if a jury con- victed whenever the evidence generated a similar degree of belief in guilt, the verdicts in this group of cases would tend to be right about half the time." Finkelstein & Fairley 504. *See also* Broun & Kelly, *supra* note 5, at 31; Kaplan 1073. But the functional relationship between subjective probabilities and likely outcomes is far more complex than this equation assumes, for it turns on such factors as how easy or difficult it is for either party to generate a given level of belief in a false propositon. *See* the related discussion at p. 1385 *infra*.

[64] *See* p. 1340 *supra*.

[65] *See* note 37 *supra*. Indeed, some statistical evidence, *see, e.g.,* note 48 *supra,* is so general and remote from the particular case as to be of only marginal rele- vance — if that. Of course, if plaintiff can satisfactorily account for the evidentiary omission, the statistical evidence alone might well suffice. *See* note 102 *infra*. And what constitutes a satisfactory explanation of the failure to adduce non-statistical evidence might *itself* turn, at least in part, on the level of probability suggested by the statistics. This possibility was not noted by Ball, *supra* note 37, at 822-823.

[66] *See id.* at 823. This seems to me the only sensible meaning that can be at- tached to such pronouncements as that "a verdict must be based upon what the jury finds to be facts rather than what they find to be more 'probable'." Lampe v. Franklin Am. Trust Co., 339 Mo. 361, 384, 96 S.W.2d 710, 723 (1936). *Ac- cord,* Frazier v. Frazier, 228 S.C. 149, 168, 89 S.E.2d 225, 235 (1955). *See also* Note, *Variable Verbalistics — the Measure of Persuasion in Tennessee,* 11 VAND. L. REV. 1413 (1958).

the blue buses, however careful, would have to pay for *five*-fifths of all unexplained blue bus accidents — a result as inefficient as it is unfair.[67]

A fortiori, when the governing standard of proof is more stringent still, the mathematics taken alone would typically fall short of satisfying it. As the California Supreme Court put it in the *Collins* case, "no mathematical equation can prove beyond a reasonable doubt (1) that the guilty [party] *in fact* possessed the characteristics described by the People's witnesses, or even (2) that only *one* [party] possessing those characteristics could be found in the [relevant] area."[68]

But the fact that mathematical evidence *taken alone* can rarely, if ever, establish the crucial proposition with sufficient certitude to meet the applicable standard of proof does not imply that such evidence — *when properly combined with other, more conventional, evidence in the same case* — cannot supply a useful link in the process of proof. Few categories of evidence indeed could ever be ruled admissible if each category had to stand on its own, unaided by the process of cumulating information that characterizes the way any rational person uses evidence to reach conclusions. The real issue is whether there is any acceptable way of combining mathematical with non-mathematical evidence. If there is, mathematical evidence can indeed assume the role traditionally played by other forms of proof. The difficulty, as we shall see, lies precisely in this direction — in the discovery of an acceptable integration of mathematics into the trial process. I now turn to a consideration of the only plausible mode of integration yet proposed.

D. A Possible Solution

In deciding a disputed proposition, a rational factfinder probably begins with some initial, a priori estimate of the likelihood of the proposition's truth, then updates his prior estimate in light of discoverable evidence bearing on that proposition, and arrives finally at a modified assessment of the proposition's likely truth in light of whatever evidence he has considered. When many items of evidence are involved, each has the effect of adjusting, in greater or lesser degree, the factfinder's evaluation of the probability that the proposition before him is true. If this

---

[67] There are, of course, possibilities of proportioned verdicts or other forms of judicial compromise in obviously doubtful cases, but these are not without their own subtle difficulties. *See* Allen, Coons, Freund, Fuller, Jones, Kaufman, Nathanson, Noonan, Ruder, Schuyler, Sowle & Snyder, *On Approaches to Court Imposed Compromises — The Uses of Doubt and Reason*, 58 Nw. U.L. Rev. 750, 795 (1964).

[68] People v. Collins, 68 Cal. 2d 319, 330, 438 P.2d 33, 40, 66 Cal. Rptr. 497, 504 (1968).

incremental process of cumulating evidence could be given quantitative expression, the factfinder might then be able to combine mathematical and non-mathematical evidence in a perfectly natural way, giving each neither more nor less weight than it logically deserves.

A quantitative description of the ordinary process of weighing evidence has long been available.[69] Before deciding whether that description can be put to the suggested use of enabling the factfinder to integrate mathematical and non-mathematical evidence, it will be necessary to develop the description briefly here.

Suppose X represents a disputed factual proposition; that is, the question for the trier is whether X is true or false. And suppose E represents some other proposition, the truth of which has just been established. Prior to learning E, the trier's subjective probability assessment of X was $P(X)$. After learning E, the trier's assessment of X will typically change. That is, the trier's subjective probability for X *given the fact that E is true*, designated $P(X|E)$,[70] will ordinarily differ from the trier's prior subjective probability for X.[71] The problem is to determine exactly how $P(X|E)$, the probability of X given E, can be calculated in terms of $P(X)$ and such other quantities as are available — to discover, that is, how the receipt of evidence E quantitatively transforms $P(X)$ into $P(X|E)$.

The solution to that problem, commonly known as Bayes' Theorem, will be summarized verbally after its mathematical formulation has been explained. The theorem follows directly from two elementary formulas of probability theory: if A and B are any two propositions, then:

(1) $P(A \& B) = P(A|B) \cdot P(B)$

(2) $P(A) = P(A \& B) + P(A \& \text{not-}B)$.[72]

---

[69] Reverend Thomas Bayes, in *An Essay Toward Solving a Problem in the Doctrine of Chance*, PHILOSOPHICAL TRANS. OF THE ROYAL SOCIETY (1763), suggested that probability judgments based on intuitive guesses should be combined with probabilities based on frequencies by the use of what has come to be known as Bayes' Theorem, a fairly simple formula that is derived at p. 1352 *infra*. More recently, it has become common to think of Bayes' Theorem as providing "a quantitative description of the ordinary process of weighing evidence." I. GOOD, PROBABILITY AND THE WEIGHING OF EVIDENCE 62 (1950). *See also* J. VENN, THE LOGIC OF CHANCE ch. 16–17 (3rd ed. 1868).

[70] $P(X|E)$ is usually read "P of X given E," or "the probability of X given the truth of E."

[71] A conditional probability like $P(X|E)$ is often understood to assume the given condition E as a certainty. One can as readily interpret $P(X|E)$, however, as measuring the degree to which the trier would believe X *if* he were sure of E. *See* note 63 *supra*.

[72] To make these formulas intuitively transparent, consider exactly what they assert. The first asserts that the probability that A and B are *both* true equals

These formulas can be shown to imply

$$(3) \quad P(X|E) = \left[\frac{P(E|X)}{P(E)} \right] \cdot P(X)$$

and

$$(4) \quad P(E) = P(E|X) \cdot P(X) + P(E|\text{not-}X) \cdot P(\text{not-}X).^{73}$$

And, using (4) to calculate $P(E)$ in (3), we obtain

$$(5) \quad P(X|E) =$$

$$\left[\frac{P(E|X)}{P(E|X) \cdot P(X) + P(E|\text{not-}X) \cdot P(\text{not-}X)} \right] \cdot P(X).$$

Formula (5), known as Bayes' Theorem, determines $P(X|E)$ in terms of $P(X)$, $P(E|X)$, and $P(E|\text{not-}X)$.[74] In the abbreviated form of formula (3), Bayes' Theorem expresses the common sense notion that, to obtain $P(X|E)$ from $P(X)$, one multiplies the latter by a factor representing the probative force of E — that is, a factor equal to the ratio of $P(E|X)$ (designating the probability of E *if* X is true) to $P(E)$ (designating the probability of E whether or *not* X is true).[75]

---

the product of two other probabilities: the probability that B is true, multiplied by the probability that A would also be true *if* B were true. *See* notes 63 & 71 *supra*.

The second formula asserts that the probability that A is true equals the sum of two other probabilities: the probability that A is true and B is *true*, plus the probability that A is true and B is *false*. Of course, B is either true or false, mutually exclusive possibilities, so the second formula reduces to the assertion that the probability that one of two mutually exclusive events will occur equals the sum of the probabilities of each event's occurrence. For example, if a well-shuffled deck contains ten white cards, ten gray cards, and eighty black cards, the probability that a card chosen randomly from the deck will be either white or gray equals .2, the sum of the probability that it will be white (.1) and the probability that it will be gray (.1).

[73] Formula (1) implies that $P(E\&X) = P(E|X) \cdot P(X)$. But E&X is identical with X&E so $P(E\&X) = P(X\&E) = P(X|E) \cdot P(E)$. Thus $P(X|E) \cdot P(E) = P(E|X) \cdot P(X)$, from which we obtain formula (3) by dividing $P(E)$ into both sides of this equation. Formula (2) implies that $P(E) = P(E\&X) + P(E \& \text{not-}X)$. Applying formula (1), we know that $P(E\&X) = P(E|X) \cdot P(X)$ and that $P(E \& \text{not-}X) = P(E|\text{not-}X) \cdot P(\text{not-}X)$, from which we obtain formula (4) by adding these two terms.

[74] The only other variable in (5), $P(\text{not-}X)$, is equal to $1 - P(X)$.

[75] Another formulation of Bayes' Theorem, less conventional and for most purposes less convenient, might nonetheless be noted here inasmuch as it may be easier to grasp intuitively. If $O(X)$ represents the "odds of X," defined as $P(X)/P(\text{not-}X)$, then

$$O(X|E) = \frac{P(E|X)}{P(E|\text{not-}X)} \cdot O(X),$$

which in effect defines the "probative force" of E with respect to X, written $PF(E \text{ wrt } X)$, as the ratio of the probability that E would be true if X were *true*

Perhaps the easiest way to express Bayes' Theorem for the non-mathematician, though it is not the most convenient expression for actual use of the theorem, is to say that

$$(6) \quad P(X|E) = \frac{P(E \, \& \, X)}{P(E)} = \frac{P(E|X) \cdot P(X)}{P(E)} \quad .$$

This simply asserts that the probability of X being true if E is known to be true, designated $P(X|E)$, may be determined by measuring how often, out of all cases in which E is true, will X *also* be true — that is, by calculating the ratio of $P(E \, \& \, X)$ to $P(E)$. That ratio, in turn, equals $P(E|X) \cdot P(X)$ divided by $P(E)$, which completes the equation in Bayes' Theorem.

To give a concrete example, let X represent the proposition that the defendant in a particular murder case is guilty, and let E represent the evidentiary fact that the defendant left town on the first available plane after the murder was committed. Suppose that, prior to learning E, an individual juror would have guessed, on the basis of all the information then available to him, that the defendant was approximately twice as likely to be guilty as he was to be innocent, so that the juror's prior subjective probability for the defendant's guilt was $P(X) = 2/3$, and his prior subjective probability for the defendant's innocence was $P(\text{not-}X) = 1/3$. What effect should learning E have upon his probability assessment — *i.e.*, $P(X|E)$? The answer to that question will depend, of course, on how much more likely he thinks a guilty man would be to fly out of town immediately after the murder than an innocent man would be.

Suppose his best guess is that the probability of such flight if the defendant is guilty, designated $P(E|X)$, is twenty percent, and that the probability of such flight if the defendant is innocent, designated $P(E|\text{not-}X)$, is ten percent. Then $P(E|X) = 1/5$ and $P(E|\text{not-}X) = 1/10$. Recall formula (4):

$$P(E) = P(E|X) \cdot P(X) + P(E|\text{not-}X) \cdot P(\text{not-}X).$$

---

to the probability that E would be true if X were *false*. It is tempting, but incorrect, to assume that $PF(E_1 \, \& \, E_2 \, \text{wrt} \, X)$, the probative force of the combination of E_1 and E_2 with respect to X, always equals the product of their separate probative forces, $PF(E_1 \, \text{wrt} \, X) \cdot PF(E_2 \, \text{wrt} \, X)$. If E_1 and E_2 are not *conditionally independent* of both X and not-X (*i.e.*, if $P(E_1 \, \& \, E_2|X) \neq P(E_1|X) \cdot P(E_2|X)$ *or* if $P(E_1 \, \& \, E_2|\text{not-}X) \neq P(E_1|\text{not-}X) \cdot P(E_2|\text{not-}X)$), then one *cannot conclude* that
$$O(X|E_1 \, \& \, E_2) = PF(E_1 \, \text{wrt} \, X) \cdot PF(E_2 \, \text{wrt} \, X) \cdot O(X),$$
although the above formula *does* hold if conditional independence obtains (*i.e.*, if $P(E_1 \, \& \, E_2|X) = P(E_1|X) \cdot P(E_2|X)$ *and* $P(E_1 \, \& \, E_2|\text{not-}X) = P(E_1|\text{not-}X) \cdot P(E_2|\text{not-}X)$). *See* pp. 1366–68 *infra*. These difficulties are overlooked in Kaplan 1085–86.

As applied to this case, we have

$$P(E) = (1/5)(2/3) + (1/10)(1/3) = 1/6.$$

In other words, given his prior assessment of $P(X) = 2/3$, the juror's best estimate of the probability of the defendant's flight, $P(E)$, would have been $1/6$. But if he knew that the defendant were in fact guilty, his best estimate of the probability that the defendant would flee, $P(E|X)$, would be $1/5$. Learning that he actually did flee should thus multiply the juror's prior assessment by the ratio

$$\frac{P(E|X)}{P(E)} = \frac{1/5}{1/6} = \frac{6}{5}.$$

Applying formula (3),

$$P(X|E) = \left[\frac{P(E|X)}{P(E)}\right] \cdot P(X),$$

or

$$P(X|E) = \left[\frac{6}{5}\right] \cdot \frac{2}{3} = \frac{4}{5}.$$

Therefore, his subsequent probability assessment of the defendant's guilt, after learning of his flight, should be $4/5$. The evidence of flight should thus increase the juror's subjective probability estimate of the defendant's guilt from $2/3$ to $4/5$ — assuming that he thinks there would be a $1/5$ probability of flight if the defendant were guilty and a $1/10$ probability of flight if he were not.

Given this precise a tool for cumulating evidence in a quantitative way, there might seem to be no obstacle to assimilating mathematical evidence into the trial process. Indeed, two commentators — one a lawyer, the other a statistician — have proposed doing exactly that. In an article in this *Review*,[76] Michael Finkelstein and William Fairley recently suggested that mathematically expert witnesses might be employed to explain to jurors the precise probative force of mathematical evidence in terms of its quantitative impact on the jurors' prior probability assessments.[77] As the next section of this article tries to show, although their analysis is both intriguing and illuminating, neither the technique proposed by Finkelstein and Fairley, nor any other

---

[76] Finkelstein & Fairley, *supra* note 5.

[77] *Id.* 502, 516–17. Although he is somewhat less precise about his suggestion, another commentator may have intended to advance a similar proposal in an earlier article. *See* Cullison, *supra* note 23, at 505. And a roughly equivalent proposal was in fact put forth over twenty years ago. *See* I. Good, Probability and the Weighing of Evidence 66–67 (1950).

like it, can serve the intended purpose at an acceptable cost. It will be necessary first, however, to review the method they have suggested. To that end, it is useful to begin with the hypothetical case Finkelstein and Fairley posit.

A woman's body is found in a ditch in an urban area. There is evidence that the deceased quarreled violently with her boy-friend the night before and that he struck her on other occasions. A palm print similar to the defendant's is found on the knife that was used to kill the woman. Because the information in the print is limited, an expert can say only that such prints appear in no more than one case in a thousand. The question Finkelstein and Fairley ask themselves is how the jury might best be informed of the precise incriminating significance of that finding.

By itself, of course, the "one-in-a-thousand" statistic is not a very meaningful one. It does not, as the California Supreme Court in *Collins* showed,[78] measure the probability of the defendant's innocence — although many jurors would be hard-pressed to understand why not. As Finkelstein and Fairley recognize,[79] even if there were as few as one hundred thousand potential suspects, one would expect approximately one hundred persons to have such prints; if there were a million potential suspects, one would expect to find a thousand or so similar prints. Thus the palm print would hardly pinpoint the defendant in any unique way.

To be sure, the finding of so relatively rare a print which matches the defendant's is an event of significant probative value, an event of which the jury should almost certainly be informed.[80] Yet the *numerical index* of the print's rarity, as measured by the frequency of its random occurrence, may be more misleading than enlightening, and the jury should be informed of that frequency — if at all — only if it is also given a careful explanation that there might well be many other individuals with similar prints. The jury should thus be made to understand that the frequency figure does not in any sense measure the probability of the defendant's innocence.[81]

---

[78] 68 Cal. 2d at 320, 438 P.2d at 33, 66 Cal. Rptr. at 497.

[79] Finkelstein & Fairley 497.

[80] Contrary to the implication in *id.*, nothing whatever in the *Collins* opinion suggests that the palm-print evidence *itself* would be thought to have insufficient probative value to be admissible. Had that been the view of the *Collins* court, it would have been forced to conclude that the evidence of the six matching characteristics was inadmissible on the facts of that case. Of course, the court concluded no such thing, and rejected only the prosecutor's particular attempt to *quantify* the probative force of the coincidence of characteristics. *Cf.* note 2 *supra*.

[81] *See* pp. 1336–38 *supra*.

Finkelstein and Fairley are distressed that this might leave the jury with too little information about the print's full probative value. The solution they propose to meet this difficulty is the use at trial of Bayes' Theorem; it is this solution to which I particularly object.[82] Let X represent the proposition that the defendant used the knife to kill his girlfriend, and let E represent the proposition that a palm print resembling the defendant's was found on the knife that killed her. $P(E|X)$ is the probability of finding a palm print resembling the defendant's on the murder weapon if he was in fact the one who used the knife to kill, and $P(E|not\text{-}X)$ is the probability of finding a palm print resembling the defendant's on the murder weapon if he was *not* the knife-user. $P(X)$ represents the trier's probability assessment of the truth of X *before* learning E, and $P(X|E)$ represents the trier's probability assessment of the truth of X *after* learning E. Finally, $P(not\text{-}X)$ represents the trier's probability assessment of the falsity of X before learning E, so that $P(not\text{-}X) = 1-P(X)$. Recall now Bayes' Theorem:

$$P(X|E) = \left[\frac{P(E|X)}{P(E|X) \cdot P(X) + P(E|not\text{-}X) \cdot P(not\text{-}X)} \right] \cdot P(X).$$

In applying this formula, Finkelstein and Fairley "assume for simplicity that defendant would inevitably leave such a print," [83] so that $P(E|X) = 1$. They also state that the probability $P(E|not\text{-}X)$ equals "the frequency of the print in the suspect population." [84] In other words, they assume that the probability of finding a print like the defendant's on the knife, if the defendant did not in fact use the knife to kill his girlfriend, is equal to the probability that a randomly chosen person would have a print like the defendant's. In a later section of this article,[85] I will try to show that both of those assumptions are entirely unrealistic, that this error substantially distorts the results derived by Finkelstein and Fairley, and — most importantly — that the error reflects not so much carelessness in their application of math-

---

[82] In fairness, it should be said that the solution is put forth quite tentatively, *see* Finkelstein & Fairley 502, and that, although Finkelstein and Fairley do expressly advocate its adoption, *see id.* 516–17, the main thrust of their work is to enlarge our understanding of the process of evidentiary inference through the application of Bayesian techniques, a task that they perform admirably and with a candor and explicitness that makes it possible for others to criticize and build upon their initial efforts. My disagreement is only with their suggested use of those techniques, however improved, *at trial*.

[83] *Id.* 498.

[84] *Id.* 500.

[85] *See* pp. 1362–65 *infra*.

ematical methods as an inherent bias generated by the use of the methods themselves in the trial process.[86]

For now, however, my only purpose is to see where the method, as Finkelstein and Fairley apply it, leads us. They undertake, using Bayes' Theorem, to construct a table showing the resulting value of $P(X|E)$ for a range of prior probabilities $P(X)$ varying from .01 to .75, and for a range of different values for the frequency of the print in the suspect population varying from .001 to .50. Nine typical values for $P(X|E)$ taken from the table Finkelstein and Fairley obtain (using the two simplifying assumptions noted above) are as follows:[87]

TABLE

Posterior Probability of X Given E as a Function
of Frequency of Print and Prior Probability

Frequency of Print	Prior Probability, $P(X)$		
	.01	.25	.75
.50	.019	.400	.857
.10	.091	.769	.967
.001	.909	.997	.9996

The table shows, for example, that if a print like defendant's occurs with a frequency of one in a thousand, and if the trier's prior assessment of the probability that defendant used the knife to kill his girlfriend is one in four before he learns of the palm-print evidence E, then the palm-print evidence should increase to .997 the trier's posterior assessment of the probability that defendant used the knife to kill: $P(X|E) = .997$.

Finkelstein and Fairley would first have each juror listen to the evidence and arrive at a value for $P(X)$ based upon his own view of the non-mathematical evidence (in this case, the prior quarrels and violent incidents). Then an expert witness would in effect show the jury the appropriate row from a table like the above, choosing the row to correspond with the testimony as to the print's frequency, so that each juror could locate the appro-

---

[86] I do not argue that the methods in question will invariably display any intrinsic bias *outside* the trial context, at least insofar as their use can be subjected to continuing scrutiny and improvement over time. In the trial setting, however, there are institutional goals and constraints that effectively preclude the undistorted use of mathematical techniques. *See generally* pp. 1358–77 *infra*.

[87] Finkelstein & Fairley 500.

priate value of $P(X|E)$ as his final estimate of the probability that the defendant was in fact the knife-user. If the print's frequency were established to be .001, for example, the jurors need only be shown the last row; if it were .10, the jurors need only be shown the second row.[88] In this way, Finkelstein and Fairley argue, the frequency statistic would be translated for the jury into a probability statement which accurately describes its probative force. And, the authors add,[89] most of the respondents in an informal survey conducted by them would have derived higher final probabilities by this method than they did without the assistance of Bayes' Theorem. "Probably the greatest danger to a defendant from Bayesian methods," the authors conclude, "is that jurors may be surprised at the strength of the inference of guilt flowing from the combination of their prior suspicions and the statistical evidence. But this, if the suspicions are correctly estimated, is no more than the evidence deserves." [90]

Is it? We will be in a better position to answer that question at the end of the next section, which examines the costs we must be prepared to incur if we would follow the path Finkelstein and Fairley propose. What will presently be identified as certain costs of quantified methods of proof might conceivably be worth incurring if the benefit in increased trial accuracy were great enough. It turns out, however, that mathematical proof, far from providing any clear benefit, may in fact decrease the likelihood of accurate outcomes. It is the accuracy issue that I will consider first.

E. The Costs of Precision

1. The Distortion of Outcomes. — a. The Elusive Starting Point. — It is of course necessary, if the trier is to make any use at all of techniques like that proposed by Finkelstein and Fairley, for him first to settle on a numerical value for $P(X)$, his assessment of the probability of X prior to the mathematical bombshell typically represented by the evidence E. But the lay trier will surely find it difficult at best, and sometimes impossible, to attach to $P(X)$ a number that correctly represents his real prior assessment. Few laymen have had experience with the assignment of probabilities, and it might end up being a matter of pure chance whether a particular juror converts his mental state of partial certainty to a figure like .33, .43, or somewhere in between. An estimate of .5 might signify for one juror a guess

---

[88] *Id.* 502. If the print's frequency were disputed, the jurors would of course have to be shown more than one row of such a table.
[89] *Id.* 502–03 n.33.
[90] *Id.* 517.

in the absence of any information and, for another, the conclusion of a search that has narrowed the inquiry to two equally probable suspects. And a juror's statement that he is "four-fifths sure," to revert to an earlier example,[91] is likely, in all but the simplest cases, to be spuriously exact.

Because the Finkelstein-Fairley technique thus compels the jury to begin with a number of the most dubious value, the use of that technique at trial would be very likely to yield wholly inaccurate, and misleadingly precise, conclusions. Even setting this threshold problem aside, major difficulties remain.

   b. The Case of the Mathematical Prior. — Finkelstein and Fairley consider the application of their technique primarily to cases in which the prior probability assessment of a disputed proposition is based on non-mathematical evidence [92] and is then modified by the application of Bayes' Theorem to some further' item of evidence that links the defendant to the case in a quantifiable way [93] or sheds some quantifiable light upon his conduct.[94] When statistical evidence is so used to modify a prior probability assessment, it is true, as the authors claim, that

> Bayesian analysis would demonstrate that the evidentiary weight of an impressive figure like one in a thousand — which might otherwise exercise an undue influence — would depend on the other evidence in the case, and might well be relatively insignificant if the prior suspicion were sufficiently weak.[95]

What they ignore, however, is that in most cases, whether civil or criminal, it will be none other than this "impressive figure like one in a thousand" that their general approach to proof would highlight. For, in most cases, the mathematical evidence will not

---

   [91] *See* p. 1346 *supra.*
   [92] *See* Finkelstein & Fairley 498–505. The authors do assert that "[u]nder certain restricted conditions, useful prior probabilities can be estimated on the basis of objective population statistics without resort to subjective evaluations," *id.* 506, and discuss studies of the use of statistics to determine prior probabilities in Polish paternity suits. *Id.* 506–09. But while discussing others' suggestions of the uses judges could make of such statistically based prior probability assessments, Finkelstein and Fairley do not consider the application of their technique of using Bayesian analysis at trial to such prior probabilities, nor consider its implications. Moreover, the authors suggest that, even if statistics could be used to arrive at an objective prior probability, "[w]here in [the judge's] opinion the facts showed that the case was either stronger or weaker than usual, he could subjectively adjust the prior [probability] accordingly." *Id.* 509.
   [93] *E.g.*, the palm-print hypothetical, *discussed at* pp. 1355–58 *supra*, or the *Collins* case, *discussed at* pp. 1334–37 *supra*, or the Thunderbird hypothetical, *discussed at* p. 1342 *supra.*
   [94] *E.g.*, the parking hypothetical, *discussed at* p. 1340 *supra.*
   [95] Finkelstein & Fairley 517.

be such as to modify a prior probability assessment by furnishing added data about the specific case at hand. Instead, the mathematical evidence will typically bear only upon the broad category of cases of which the one being litigated will be merely an instance. Recall, for example, the situation in which sixty percent of all barrel-falling incidents were negligently caused,[96] the situations in which ninety-eight percent of all heroin was illegally imported,[97] the case in which four out of five blue buses belonged to the defendant,[98] or the prosecution in which the mistress was the murderess in ninety-five percent of all known similar instances.[99] In all of these cases, the mathematical evidence E simply describes the *class* of cases of which the litigated case is one. In such cases, E can *only* shed light on what initial value to assign to $P(X)$.[100] Thus, the statistical information in these cases will, if given to the jury, create a high probability assessment of civil or criminal liability — and there is no assurance that the jury, either with or without the aid of Bayes' Theorem, will be able to make all of the adjustments in that high prior assessment that will be called for by the other evidence (or lack of it) that the rest of the trial reveals. The problem of the overpowering number, that one hard piece of information, is that it may dwarf all efforts to put it into perspective with more impressionistic sorts of evidence. This problem of acceptably combining the mathematical with the non-mathematical evidence is not touched in these cases by the Bayesian approach.

In situations of the sort being examined here, however, when the thrust of the mathematical evidence is to shed light on the probability assessment with which the trier ought rationally to begin, there is at least one way to take the evidence into account at trial without incurring the risk that the jury will give it too much weight when undertaking to combine the mathematical

---

[96] *See* p. 1339 *supra.*

[97] *See* pp. 1339–40 & p. 1343 *supra.*

[98] *See* pp. 1340–41 *supra.*

[99] *See* p. 1341 *supra.*

[100] When the mathematical evidence E simply describes the *class* of cases of which the litigated case is one, the truth or falsity of the litigated proposition X in no way affects the probability of the truth of E; *i.e.*, $P(E|X)$ equals $P(E)$. Bayes' Theorem then simply asserts that $P(X|E) = P(X)$. *See* equation (3), p. 1352 *supra.* Hence, in all such cases, E can suggest what *initial* value to assign to $P(X)$ but cannot serve to *refine* that initial value — as can, for example, evidence of the palm-print variety. This formulation makes more precise the common-sense notion, *cf.* p. 1346 *supra,* that the sort of statistical evidence that was offered in the bus case pertains not to the *particular* dispute being litigated but to a broad *category* of possible disputes. Evidence E is of this character if, although E is *relevant* to the disputed proposition X, it is nonetheless true that $P(E|X) = P(E)$.

datum with fuzzier information. Let the judge rather than the jury weigh the probabilistic proof in order to determine whether it might not be both equitable and conducive to an accurate out-come to shift to the other side the burden of producing some be-lievable evidence to take the case outside the general rule seem-ingly established by the probabilities.[101] If one is to avoid a distortion in results, however, any such proposal must be quali-fied, at least when the question is one of the defendant's identity, by the principle that a party is not entitled to a jury verdict on statistical evidence alone absent some plausible explanation for his failure to adduce proof of a more individualized character.[102]

But the difficulty that calls forth this solution is not limited to cases in which the mathematics points to a readily quantifiable prior assessment. The problem — that of the overbearing im-pressiveness of numbers — pervades all cases in which the trial use of mathematics is proposed. And, whenever such use is in fact accomplished by methods resembling those of Finkelstein and Fairley, the problem becomes acute.

c. The Dwarfing of Soft Variables. — The syndrome is a familiar one: If you can't count it, it doesn't exist. Equipped with a mathematically powerful intellectual machine, even the

[101] For example, in line with the suggested approach, the judge might decide to employ the doctrine of *res ipsa loquitur, see* note 33 *supra,* or any of a variety of rebuttable presumptions. *See, e.g.,* O'Dea v. Amodeo, 118 Conn. 58, 170 A. 486 (1934) (presumption of father's consent to son's operation of automobile); Hinds v. John Hancock Mut. Life Ins. Co., 155 Me. 349, 354-67, 155 A.2d 721, 725-32 (1959) (presumption against suicide). One of the traditional functions of the use of presumptions, at least those rebuttable by any substantial contrary evidence, is "to make more likely a finding in accord with the balance of probabili-ty." Morgan, *Instructing the Jury upon Presumptions and Burden of Proof,* 47 HARV. L. REV. 59, 77 (1933).

[102] *See* p. 1349 *supra.* If the statistical evidence standing alone establishes a sufficiently high prior probability of X, and a satisfactory explanation is pro-vided for the failure to adduce more individualized proof, there seems no defensible alternative (absent believable evidence contrary to X) to directing a verdict for the party claiming X, for no factual question remains about which the jury can reason, and directing a verdict the other way would be more likely to lead to an unjust result. If, however, more individualized proof *is* adduced, and if the party opposing X has discharged the burden (created by the statistical evidence, *see* note 33 *supra*) of producing believable evidence to the contrary, the question remains whether the risk of distortion created by informing the trier of fact of the potentially overbearing statistics so outweighs the probative value of such statistics as to compel their judicial exclusion. If this situation arises in a criminal case, *see, e.g.,* the heroin hypotheticals, p. 1340 & p. 1343 *supra;* the police hypothetical, p. 1341 *supra;* and the mistress hypothetical, *id.,* the added threats to important values, *see* pp. 1368-75 *infra,* should probably suffice, in combination with the danger of a distorted outcome, to outweigh the probative value of the statistics. But if the situation arises in a civil case, as in the barrel hypothetical, p. 1339 *supra,* or in the bus hypothetical, p. 1340 *supra,* all that I am now prepared to say is that the question of admissibility seems to me a very close one.

most sophisticated user is subject to an overwhelming temptation to feed his pet the food it can most comfortably digest. Readily quantifiable factors are easier to process — and hence more likely to be recognized and then reflected in the outcome — than are factors that resist ready quantification. The result, despite what turns out to be a spurious appearance of accuracy and complete-ness, is likely to be significantly warped and hence highly suspect.

The best illustration is none other than the computations per-formed by Finkelstein and Fairley themselves in their palm-print hypothetical. To begin with, they assume that, if the defendant had in fact used the knife to kill his girlfriend, then a palm print resembling his would certainly have been found on it, *i.e.*, $P(E|X) = 1$.[103] Had they not been moved by the greater ease of applying Bayes' Theorem under that assumption, the authors would surely have noted that a man about to commit murder with a knife might well choose to wear gloves, or that one who has committed murder might wipe off such prints as he happened to leave behind. Thus $P(E|X)$ equals not 1, but 1–g, where g represents the combined probability of these contingencies and of the further contingency, later considered by Finkelstein and Fairley,[104] that such factors as variations within the suspected source [105] might prevent a print left by the defendant from seem-ing to match his palm.[106]

Far more significantly, Finkelstein and Fairley equate the frequency of the palm print in the suspect population with $P(E|not\text{-}X)$, the probability of finding a print like the defend-ant's on the knife if he did not use it to kill.[107] That equation, however, strangely assumes that finding an innocent man's palm print on the murder weapon must represent a simple coincidence. If that were so, then the likelihood of such a coincidence, as measured by the print's frequency in the population, would of course yield the probability that a print like the defendant's would appear on the knife despite his innocence.[108] But this ig-nores the obvious fact that the print might have belonged to the defendant after all — without his having used the knife to kill the girl. He could simply have been framed, the real murderer having worn gloves when planting the defendant's knife at the

---

[103] Finkelstein & Fairley 498.

[104] *Id.* 509–11.

[105] The same palm might leave a variety of seemingly different partial prints.

[106] Significantly, the authors take account of the problem of source variations, Finkelstein & Fairley 510, but neglect to consider the other, less readily quanti-fiable, components of the variable g. *See id.* 498 n.22.

[107] *Id.* 498, 500.

[108] Even this is somewhat oversimplified since it neglects to multiply the fre-quency by the factor (1–g). *See also* note 113 *infra*.

scene of the crime. The California Supreme Court recognized that sort of possibility in *Collins*,[109] noting that a jury traditionally weighs such risks in assessing the probative value of trial testimony. Finkelstein and Fairley, however, overlook the risk of frame-up altogether [110] — despite the nasty fact that the most inculpatory item of evidence may be the item most likely to be used to frame an innocent man.[111]

One can only surmise that it was the awkwardness of fitting the frame-up possibility into their formula that blinded even these sophisticated authors, one a legal scholar and the other a teacher of statistical theory, to a risk which they could not otherwise have failed to perceive. And if *they* were seduced by the mathematical machinery, one is entitled to doubt the efficacy of even the adversary process as a corrective to the jury's natural tendency to be similarly distracted.[112]

As it turns out, the frame-up risk would have been awkward indeed to work into the calculation, increasing $P(E|\text{not-}X)$ from a value equal to the frequency of the palm print, hereinafter designated f, to a value equal to $f + F$, where F represents the

---

[109] *See* p. 1336 *supra.*

[110] The frame-up possibility was also overlooked by Broun & Kelly, *supra* note 5, at 27–28 & n.20. Finkelstein and Fairley seem to have overlooked this possibility literally by definition. They define X (the authors labeled it G; *see* pp. 1365–66 *infra*) to be "the event that defendant used the knife," apparently meaning that the defendant used the knife to kill. Finkelstein & Fairley 498. Yet they define not-X to be the event "that a palm print [was] left by someone other than the defendant," *id.*, leaving the frame-up case as one where both X and not-X are untrue, a logical impossibility.

Of course, it is possible that the authors intentionally prevented the frame-up possibility from affecting their calculations by the use of the definition of X, "the event that defendant used the knife," to mean the event that the print was the defendant's — in which case the frame-up case is included within X. But this would require another application of Bayes' Theorem to determine the probability of guilt from the knowledge of the probability that the print was the defendant's, a step that Finkelstein and Fairley clearly did not intend. Instead, they repeatedly referred to $P(X)$ as the "prior probability of guilt." *See, e.g., id.* 500; *see* pp. 1365–66 *infra.*

[111] A quite distinct problem, more serious in cases relying on human identification than in cases where identification is based on physical evidence, is that the characteristics of the defendant relied upon in the probabilistic formula may not in fact have been shared by the actually guilty individual or individuals because of some mistake in observation or memory. Such risks of error, like the risk of frame-up, are hard to quantify and hence likely to be underemphasized in a quantitative analysis, but differ from the risk of frame-up in that they need not perversely increase as the apparent probative value of the evidence increases.

[112] *Cf.* pp. 1337–38 *supra.* This is not to say, of course, that a professional decisionmaking body, possessed of both the skills and the time to refine its sophistication in mathematical techniques, could not overcome such a tendency — but the jury is not such a body, and making it into one for all but a very limited set of purposes would entail staggeringly high costs.

probability of frame-up.[113] Bayes' Theorem would then have assumed the messy form

$$P(X|E) = \left[\frac{(1-g)}{(1-g)\ P(X) + (f + F)\ P(not\text{-}X)} \right] \cdot P(X)\ .\ ^{114}$$

This formula may be easy enough to use when one assumes, with Finkelstein and Fairley, that $g = 0$ and $F = 0$, but is rather more troublesome to handle when one has no real idea *how* large those two probabilities are.

Moreover, it makes quite a difference to the outcome just how large the probabilities, g and f, turn out to be. Consider again the case in which the print is assumed to occur in one case in a thousand, so that $f = .001$, and in which the prior assessment is $P(X) = .25$. On those facts, Finkelstein and Fairley conclude — by treating g and F as though they were both zero — that $P(X|E) = .997$, an overwhelming probability that the defendant did the killing. If, however, we assume that $g = .1$ and $F = .1$, then the same initial values $f = .001$ and $P(X) = .25$ yield the strikingly different conclusion $P(X|E) = .75$, a very much lower probability than Finkelstein and Fairley calculated.[115]

What, then, is one to tell the jurors? That each of them should arrive somehow at his own values for g and F, as to neither of which any evidence will be available, and should then wait while a mathematician explains what $P(X|E)$ turns out to be given those values? Surely we have by now strained the system beyond its breaking point.[116] If the jurors are to work with mathematical proof in any capacity beyond that of passive ob-

---

[113] More precisely, $P(E|not\text{-}X)$ equals not simply $f+F$, but $(1-g)f+F$. *Cf.* note 108 *supra*. This fact makes the final computation even more complex, although it changes the outcome only slightly if g is small in relation to f or F.

[114] *See* equation (5), p. 1352 *supra*. The complete equation, taking into account note 113 *supra*, is more involved still:

$$P(X|E) = \left[\frac{(1-g)}{(1-g)\ P(X) + [(1-g)\ f + F]\ P(not\text{-}X)} \right] \cdot P(X).$$

[115] Indeed, if $F = .25$, $f = .1$, $g = .1$, and $P(X) = .25$, then $P(X|E) = .469$, not .769 (as Finkelstein and Fairley calculated, *id.* 500) — this time enough of a difference to cross even the "preponderance" line of .5, making $P(X|E)$ *less* likely than not rather than *more* likely than not, as the Finkelstein-Fairley method would suggest. As Professor Howard Raiffa pointed out to me upon reading this argument, it highlights how useful mathematics can be in illuminating the source, size, and structure of such distortions and underscores the point that my objection is not to mathematical analysis as such but to its formal use at trial. *See also* note 132 *infra*.

[116] *See* pp. 1375–77 *infra*. Further complicating the picture, there will often be dispute as to the truth of such underlying evidentiary facts as E which call forth the use of statistics, a circumstance that makes the use of Bayes' Theorem more complex still.

servers, they will in the end have to be given the numbers com-
puted by Finkelstein and Fairley and asked to draw their own
conclusions, keeping in mind — unless the judge who instructs
them is as blinded by the formulas as the authors were — that
possibilities such as frame-up distort the figures in the table so
that they overstate the truth by some indeterminate amount.

But then we have come full circle. At the outset some way
of integrating the mathematical evidence with the non-mathe-
matical was sought, so that the jury would not be confronted
with an impressive number that it could not intelligently com-
bine with the rest of the evidence, and to which it would there-
fore be tempted to assign disproportionate weight. At first glance,
the use Finkelstein and Fairley made of Bayes' Theorem ap-
peared to provide the needed amalgam. Yet, on closer inspection,
their method too left a number — the exaggerated and much *more*
impressive $P(X|E) = .997$ — which the jury must again be asked
to balance against such fuzzy imponderables as the risk of frame-
up or of misobservation, if indeed it is not induced to ignore
those imponderables altogether.

What is least clear in all of this is whether the proponents
of mathematical proof have made any headway at all. Even as-
suming with Finkelstein and Fairley that the accuracy of trial
outcomes could be somewhat enhanced if *all* crucial variables
could be quantified precisely and analyzed with the aid of Bayes'
Theorem, it simply does not follow that trial accuracy will be
enhanced if *some* of the important variables are quantified and
subjected to Bayesian analysis, leaving the softer ones — those
to which meaningful numbers are hardest to attach — in an im-
pressionistic limbo. On the contrary, the excessive weight that
will thereby be given to those factors that can most easily be
treated mathematically indicates that, on balance, more mistakes
may well be made with partial quantification than with no quan-
tification at all.[117]

d. Asking the Wrong Questions. — Throughout the preceding
discussion, I have referred to $P(X)$ and to $P(X|E)$, deliber-
ately eschewing the terminology employed by Finkelstein and
Fairley.[118] Instead of X, they write G — impling that $P(X|E)$

---

[117] *Cf.* Lipsey & Lancaster, *The General Theory of Second Best*, 24 REV. ECON.
STUDIES 11 (1956) (economic theory that, once some constraint prevents attain-
ment of one optimum condition, other previously optimal conditions are generally
no longer desirable as means to best solution):

> Specifically, it is *not* true that a situation in which more, but not all, of the
> optimum conditions are fulfilled is necessarily, or is even likely to be, su-
> perior to a situation in which fewer are fulfilled.

Id. 12.

[118] Finkelstein & Fairley 498–500.

represents the probability that the defendant is *guilty of murder* if a palm-print matching his is found on the murder weapon. But of course it represents no such thing, for murder means much more than causing death. To say that $P(X|E) = .997$ is to say that, given the palm-print evidence, there is a probability of .997 that the defendant used the knife to kill the deceased. It is to say nothing at all about his state of mind at the time, nothing about whether he intended to cause death, nothing about whether the act was premeditated.[119] To be sure, these elements can be called to each juror's attention, but his eyes are likely to return quickly to that imposing number on the board.

It is no accident that such matters as identity — matters that are objectively verifiable in the world outside the courtroom — lend themselves more readily to mathematical treatment than do such issues as intent — issues that correspond to no verifiable "fact" outside the verdict of the jury. It is not surprising that in none of the cases earlier enumerated under the heading of "intention" [120] was the mathematical evidence linked to any fact specifically about the defendant himself or about his own conduct or state of mind — for it is difficult even to imagine cases in which such a link could be found.

One consequence of mathematical proof, then, may be to shift the focus away from such elements as volition, knowledge, and intent, and toward such elements as identity and occurrence — for the same reason that the hard variables tend to swamp the soft. It is by no means clear that such marginal gains, if any, as we may make by finding somewhat more precise answers would not be offset by a tendency to emphasize the wrong questions.[121]

e. The Problem of Interdependence. — Essential to the application of Bayes' Theorem to derive $P(X|E)$ from $P(X)$ is that the trier be able somehow to make a prior estimate of $P(X)$, the probability of the disputed proposition X. If that estimate is arrived at after knowing or even suspecting E, then to use the information provided by E to refine the estimate through Bayes' Theorem to obtain $P(X|E)$ would obviously involve counting the same thing twice. Particularly when the proposition X goes to the identity of the person responsible for an alleged wrong, and when the tendency of E is to pinpoint the defendant as the

---

[119] Although such matters will rarely be at issue in cases where the defense rests in part on a claim of mistaken identity, it is of course possible for both identity and intent to be disputed in the same case.

[120] *See* pp. 1342–43 *supra.*

[121] This difficulty would be somewhat easier to correct than any of the others identified thus far — through the use of special instructions to the jury, or perhaps through special verdicts.

responsible person, the knowledge or suspicion of E is likely to have entered into the plaintiff's choice of this particular defend-ant. Having learned of the features of the interracial couple from the witnesses in *Collins*,[122] for example, the prosecution was hardly likely to charge someone *not* sharing those features; hav-ing found a latent palm print on the murder weapon, the State was less likely to file an indictment against a person whose palm print *failed* to match. And the trier would be hard put to dis-regard those obvious realities in attempting to derive a value for $P(X)$. The accurate application of Bayes' Theorem along the lines proposed by Finkelstein and Fairley necessarily assumes that the evidence E of quantifiable probative value can be made independent of the prior suspicion,[123] but in most trials the two will be hopelessly enmeshed.

Indeed, even if $P(X)$ is arrived at without any reliance what-ever upon E, the straightforward application of Bayes' Theorem will still entail a distorted outcome if some or all of the evidence that *did* underlie $P(X)$ was related to E, in the sense that know-ing something about the truth of X and of that underlying evi-dence would yield information one way or the other about the likely truth of E. To take a simple example, suppose that an armed robbery taking fifteen minutes to complete was committed between 3.00 a.m. and 3:30 a.m. The trier first learns E_1: that the accused was seen in a car a half-mile from the scene of the crime at 3:10 a.m. Based on this information, the trier will assess a subjective probability $P(X)$ of the accused's likely involvement in the robbery. Then the trier learns E_2: that the accused was also seen in a car a half-mile from the scene of the crime at 3:20 a.m. By itself, E_2 appears to make X, the proposition that the accused was involved, more likely than it would have seemed without any evidence as to the accused's whereabouts, so that $P(X|E_2)$ will exceed $P(X)$ if computed by applying Bayes' The-orem directly, *i.e.*, by multiplying $P(X)$ by the ratio $P(E_2|X)/P(E_2)$. Yet this is surely wrong, for if E_1 and E_2 are *both* true, then X must be false, and $P(X|E_2)$ should equal zero.[124] One

---

[122] People v. Collins, 68 Cal. 2d 319, 321, 438 P.2d 33, 34, 66 Cal. Rptr. 497, 498 (1968).

[123] Strictly speaking, the requirement is that E be *conditionally* independent of both X and not-X, where X is the proposition in dispute. *See* note 75 *supra*.

[124] An illustration of the converse situation, in which $P(X|E_1)$ and $P(X|E_2)$ are both smaller than $P(X)$ but $P(X|E_1 \& E_2)$ is much larger, is easy to construct. E_1 and E_2 could, for example, represent mutually inconsistent but nonetheless strikingly similar alibis. Independently, the alibis might seem plausible, thus re-ducing the probability of X. Taken together, not only does the inconsistency de-stroy the effectiveness of each in reducing the likelihood of X, but the similarity makes both seem contrived, allowing a more probable inference of X.

can with some effort make the appropriate adjustment — by us-
ing Bayes' Theorem to compute [125]

$$P(X|E_1 \ \& \ E_2) = \left[\frac{P(E_2|X \ \& \ E_1)}{P(E_2|E_1)} \right] \cdot P(X|E_1).$$

This means, however, that the theorem cannot be applied sequen-
tially, with one simple multiplication by $P(E|X) / P(E)$ as each
new item of evidence, E, comes in,[126] but must instead be applied
in the terribly cumbrous form shown above,[127] unless one knows
somehow that the item of evidence to which the theorem is being
applied at any given point is not linked conditionally [128] to any
evidence already reflected in one's estimate of $P(X)$. Finkelstein
and Fairley ignore that requirement; taking it into account in
order to avoid grossly inaccurate outcomes would make the ma-
chinery they propose so complex and so unwieldy that its opera-
tion, already hard enough for the juror to comprehend, would
become completely opaque to all but the trained mathematician.

   2. *The End of Innocence: A Presumption of Guilt?* — At
least in criminal cases, and perhaps also in civil cases resting on
allegations of moral fault, further difficulties lurk in the very fact
that the trier is forced by the Finkelstein-Fairley technique to
arrive at an explicit quantitative estimate of the likely truth at
or near the trial's start, or at least before some of the most
significant evidence has been put before him.[129]

   To return for a moment to the palm-print case posited by
Finkelstein and Fairley, a juror compelled to derive a quantita-
tive measure $P(X)$ of the defendant's likely guilt after having
heard no evidence at all, or at most only the evidence about the

---

[125] By successive applications of equations (1) and (3), pp. 1351–52 *supra*:

$$P(X|E_1 \ \& \ E_2) = \frac{P(X) \cdot P(E_1 \ \& \ E_2|X)}{P(E_1 \ \& \ E_2)} = \frac{P(X \ \& \ E_1 \ \& \ E_2)}{P(E_1 \ \& \ E_2)} =$$

$$\frac{P(X \ \& \ E_1) \cdot P(E_2|X \ \& \ E_1)}{P(E_1 \ \& \ E_2)} = \frac{P(E_1) \cdot P(X|E_1) \cdot P(E_2|X \ \& \ E_1)}{P(E_1) \cdot P(E_2|E_1)} =$$

$$\frac{P(E_2|X \ \& \ E_1) \cdot P(X|E_1)}{P(E_2|E_1)}.$$

[126] This limitation is simply ignored by Kaplan, *supra* note 5, at 1084–85; *see*
note 75 *supra*.

[127] More generally, $P(X|E_1 \ \& \ E_2 \ \& \ldots \& \ E_n \ \& \ E_{n+1}) =$

$$\frac{P(E_{n+1}|X \ \& \ E_1 \ \& \ldots \& \ E_n)}{P(E_{n+1}|E_1 \ \& \ldots \& \ E_n)} \cdot P(X|E_1 \ \& \ldots \& \ E_n).$$

[128] *See* note 75 *supra*.

[129] The crucial factor is not so much that of time sequence as that of priority
in thought. Even if all of the evidence has in fact been introduced before the trier
is asked to quantify the probative force of a limited part of it without taking
account of the rest, the problems discussed here still obtain.

defendant's altercations with the victim, cannot escape the task of deciding just how much weight to give the undeniable fact that the defendant is, after all, not a person chosen at random but one strongly suspected and hence accused by officials of the state after extended investigation. If the juror undertakes to assess the probative value of that fact as realistically as he can, he will have to give weight to whatever information he supposes was known to the grand jury or to the prosecutor who filed the accusation. To the extent that this supposed information contains facts that will be duplicated by later evidence, such evidence will end up being counted twice, and will thus be given more weight than it deserves.[130] And, to the extent that the information attributed to the prosecutor or the grand jury is *not* duplicative in this sense, it will include some facts that will not be, and some facts that cannot properly be, introduced at trial as evidence against the accused.[131] Inviting jurors to take account of such facts is at war with the fundamental notion that the jury should make an independent judgment based only on the evidence properly before it, and would undercut the many weighty policies that render some categories of evidence legally inadmissible.[132]

---

[130] *See* p. 1366 *supra.*

[131] Such facts would typically include, for example, hearsay of a sort that, though inadmissible at trial, may support a valid indictment. *See* Costello v. United States, 350 U.S. 359, 362–63 (1956). *But cf.* United States v. Payton, 363 F.2d 996 (2d Cir.), *cert. denied*, 385 U.S. 993 (1966).

[132] Nor would it be a satisfactory solution to instruct the jury to assign an artificially low starting value to $P(X)$ on the pretense that the accused was plucked randomly from the population, even if the jury could be trusted to follow such an instruction. Suppose that $P(X)$ would be thought to equal $1/2$ but for the suggested pretense of random selection, and suppose that the prosecution's evidence E is so compelling that, given this $P(X)$, it turns out that $P(X|E) > .999$. This would be the case, for example, if $P(E|X) = 1/2$ and $P(E|\text{not-}X) = 1/2,000$. *See* formula (5) at p. 1352 *supra*. Given this same evidence E, if the trier were to pretend that the accused had been chosen randomly from the population of the United States and were thus to treat $P(X)$ as equal to approximately $1/200,000,000$, he would obtain $P(X|E) < 1/200,000$ — a miniscule, and evidently understated, probability that the defendant did the killing. If, however, the trier could manage to treat $P(X)$ as *completely indeterminate* until at least some evidence E' (some significant part of E) had been introduced, he might rationally assign to $P(X|E')$ a value high enough to bring $P(X|E \& E')$ close to .999 or higher. *See* note 125 *supra*. What this suggests is that setting $P(X)$ equal to an artificially small (and essentially meaningless) quantity at the outset of the trial may distort the final probability downward in a way that need not occur if *no judgment at all* is made about the starting value of $P(X)$ apart from at least some significant evidence in the case. *Cf.* note 100 *supra*. But once the jury is invited to assess the probability of guilt in light of less than *all* the evidence, there can be no assurance that it will not make an initial assessment that depends on *none* of the evidence. And, as has been shown, such an assessment can cause serious distortions whether the existence of an indictment or other charge is treated as probative background information or as equivalent to random selection. It should be noted that, without using Bayes'

Moreover, even if no such problem were present, directing the jury to focus on the probative weight of an indictment or other charge, or even directing it simply to assess the probability of the accused's guilt at some point before he has presented his case, would entail a significant cost. It may be supposed that no juror would be permitted to announce publicly in mid-trial that the defendant was already burdened with, say, a sixty percent probability of guilt — but even without such a public statement it would be exceedingly difficult for the accused, for the prosecution, and ultimately for the community, to avoid the explicit recognition that, having been forced to focus on the question, the rational juror could hardly avoid reaching some such answer. And, once that recognition had become a general one, our society's traditional affirmation of the "presumption of innocence" could lose much of its value.

That presumption, as I have suggested elsewhere, "represents far more than a rule of evidence. It represents a commitment to the proposition that a man who stands accused of crime is no less entitled than his accuser to freedom and respect as an innocent member of the community." [133] In terms of tangible consequences for the accused, this commitment is significant because it can protect him from a variety of onerous restraints not needed to effectuate the interest in completing his trial; [134] because the suspension of adverse judgment that it mandates can encourage the trier to make an independent and more accurate assessment of his guilt; and because it may help to preserve an atmosphere in which his acquittal, should that be the outcome, will be taken seriously by the community. But no less important are what seem to me the intangible aspects of that commitment: its expressive and educative nature as a refusal to acknowledge prosecutorial omniscience in the face of the defendant's protest of innocence, and as an affirmation of respect for the accused — a respect expressed by the trier's willingness to listen to all the accused has to say before reaching any judgment, even a tentative one, as to his probable guilt.

It may be that most jurors would suspect, if forced to think about it, that a substantial percentage of those tried for crime

---

Theorem to compute the figures employed in this footnote, the degree of understatement that might be caused by an artificially deflated starting value for P(X) would be difficult if not impossible to assess, a fact that again serves to illustrate the usefulness of mathematical techniques in illuminating the process of proof even when, and perhaps especially when, one is rejecting the formal application of such techniques in the process itself. *See also* note 115 *supra*.

    [133] Tribe, *An Ounce of Detention: Preventive Justice in the World of John Mitchell*, 56 U. VA. L. REV. 371, 404 (1970) [hereinafter cited as Tribe].
    [134] *Id.* 404–06.

are guilty as charged. And that suspicion might find its way unconsciously into the behavior of at least some jurors sitting in judgment in criminal cases. But I very much doubt that this fact alone reduces the "presumption of innocence" to a useless fiction. The presumption retains force not as a *factual* judgment, but as a *normative* one — as a judgment that society *ought* to *speak* of accused men as innocent, and *treat* them as innocent, until they have been properly convicted after all they have to offer in their defense has been carefully weighed. The suspicion that many are *in fact* guilty need not undermine either this normative conclusion or its symbolic expression through trial procedure, so long as jurors are not compelled to articulate their prior suspicions of guilt in an explicit and precise way.

But if they *are* compelled to measure and acknowledge a factual presumption of guilt at or near each trial's start, then their underlying suspicion that such a presumption would often accord with reality may indeed frustrate the expressive and instructional values of affirming in the criminal process a normative presumption of innocence. Jurors cannot at the same time estimate probable guilt and suspend judgment until they have heard all the defendant has to say. It is here that the great virtue of mathematical rigor — its demand for precision, completeness, and candor — may become its greatest vice, for it may force jurors to articulate propositions whose truth virtually all might already suspect, but whose explicit and repeated expression may interfere with what seem to me the complex symbolic functions of trial procedure and its associated rhetoric.

To the extent that this argument and the one that immediately follows it [135] appear to run counter to rarely questioned assumptions about the transcending values of full candor and complete clarity, I should stress that the arguments in question are entirely independent of my other criticisms of mathematical methods in the trial process. Nor do I mean by advancing these arguments to suggest that departures from candor are lightly to be countenanced. Indeed, I have not proposed that anyone deceive either himself or another about the factual underpinnings of the presumption of innocence, but only that worthwhile values served by that presumption as a normative standard might be harder to secure if the probability of guilt became a matter for precise and explicit assessment and articulation early in the typical criminal trial. The point, then, is not that any factual truth should be concealed or even obscured, but only that one need not say everything all at once in order to be truthful, and that saying some things in certain ways and at certain times in the trial process

---

[135] *See* pp. 1372–75 & 1390 *infra.*

may interfere with other more important messages that the proc-
ess should seek to convey and with attitudes that it should seek
to preserve.

   3. The Quantification of Sacrifice. — This concern for the
expressive role of trial procedure is no less relevant to the trial's
end than to its start. Limiting myself here to the ordinary crim-
inal proceeding,[136] I suggest that the acceptance of anything like
the method Finkelstein and Fairley propose, given the precision
and explicitness its use demands, could dangerously undermine
yet another complex of values — the values surrounding the no-
tion that juries should convict only when guilt is beyond real
doubt.[137]

   An inescapable corollary of the proposed method, and indeed
of any method that aims to assimilate mathematical proof by
quantifying the probative force of evidence generally, is that it
leaves the trier of fact, when all is said and done, with a number
that purports to represent his assessment of the probability that
the defendant is guilty as charged.[138] Needless to say, that num-
ber will never quite equal 1.0, so the result will be to produce a
quantity — take .95 for the sake of illustration — which openly
signifies a measurable (and potentially reducible) margin of
doubt, here a margin of .05, or 1/20.

   Now it may well be, as I have argued elsewhere,[139] that there
is something intrinsically immoral about condemning a man as a
criminal while telling oneself, "I believe that there is a chance of
one in twenty that this defendant is innocent, but a 1/20 risk of
sacrificing him erroneously is one I am willing to run in the
interest of the public's —and my own — safety." It may be that

---

[136] The argument I advance here — unlike the arguments beginning at pp. 1358–
59 *supra*, and unlike at least part of the argument beginning at p. 1368 *supra*,
is meant to apply only to criminal cases, and perhaps only to those criminal cases
in which either the penalty attached or the moral blame imputed makes the crime
a sufficiently "serious" one.

[137] Although this notion is readily confused with the presumption of innocence,
discussed at pp. 1370–71 *supra*, it is in fact quite different and rests on a partially
overlapping but partially distinct set of objectives. To some extent, in fact, the
concept that conviction is proper only after all real doubt has been dispelled may
tend to undercut the purposes served by the presumption of innocence, for that
concept suggests that a defendant's acquittal signifies only the existence of some
doubt as to his guilt, whereas one function of the presumption of innocence is to
encourage the community to treat a defendant's acquittal as banishing all lingering
suspicion that he might have been guilty. *See* p. 1370 *supra.*

[138] Even when the number measures only one element of the offense and omits
an element like intent, *see* pp. 1365–66 *supra*, it sets an upper bound on the
probability of guilt, and the argument made below follows a fortiori.

[139] Tribe 385–87.

— quite apart from the particular number — there is something basically immoral in this posture, but I do not insist here on that position. All I suggest is that a useful purpose may be served by structuring a system of criminal justice so that it can avoid having to proclaim, as the Finkelstein-Fairley procedure would force us to proclaim, that it will impose its sanctions in the face of a recognized and quantitatively measured doubt in the particular case.

If the system in fact did exactly that, such compelled candor about its operation might have great value. It could generate pressure for useful procedural reform, and it should probably be considered worthwhile in itself.[140] But to let the matter rest

---

[140] Even if this were the case, I would find it difficult wholly to ignore the fact that at least some who witness the trial process might interpret a series of publicly proclaimed decisions to condemn in the face of numerically measurable doubt as teaching that the sacrifice of innocent men is not to be regarded as a terribly serious matter. *See id.* 387 n.65. Those who adopt this interpretation might become more willing than they should be to tolerate the sacrifice of others and less confident than they ought to be of their own security from unjust conviction. Although such an interpretation would be in error, it would not be *wholly* unjustified, for a society that does not recoil from confronting defendants in quantitative terms with the magnitude of its willingness to risk their erroneous conviction *is*, it seems to me, a society that takes the tragic necessity of such sacrifice less seriously than it might.

When the Supreme Court held that "the Due Process Clause protects the accused against conviction except upon proof beyond a reasonable doubt of every fact necessary to constitute the crime with which he is charged," it stressed the importance of not leaving the community "in doubt whether innocent men are being condemned," reasoning in part that such doubt would dilute "the moral force of the criminal law" and in part that it would impair the confidence of "every individual going about his ordinary affairs . . . that his government cannot adjudge him guilty of a criminal offense without convincing a proper factfinder of his guilt with utmost certainty." *In re* Winship, 397 U.S. 358, 364 (1970). If due process required less than this publicly announced insistence upon "a subjective state of certitude," *id., quoting* Dorsen & Rezneck, *In re Gault and the Future of Juvenile Law,* 1 FAMILY L.Q. 1, 26 (1967), the Court seemed to be saying, the sense of security conferred by a system that at least proclaims an unwillingness to punish in the face of palpable doubt would be irreparably eroded. *See* Tribe 389.

Both callousness and insecurity, then, might be increased by the explicit quantification of jury doubts in criminal trials — whether or not it would be *factually accurate* to describe the trial system as imposing criminal sanctions in the face of quantitatively measured uncertainty in particular cases. In considering a somewhat analogous problem in the area of accident law, Professor Calabresi has argued that there is a great difference in social cost between a set of individual market choices that indirectly sacrifice human lives by investing less than possible in life-saving resources and a collective societal choice that consciously and calculatingly sacrifices precisely the same lives for exactly the same reasons of economy. *See* Calabresi, *Reflections on Medical Experimentation in Humans,* 1969 DAEDALUS 387, 388–92. *See also* Schelling, *The Life You Save May Be Your Own,* in PROBLEMS IN PUBLIC EXPENDITURE ANALYSIS 127, 142–62 (S.B. Chase ed. 1968).

there would be wrong, for the system does *not* in fact authorize the imposition of criminal punishment when the trier recognizes a quantifiable doubt as to the defendant's guilt. Instead, the system dramatically — if imprecisely — insists upon as close an approximation to certainty as seems humanly attainable in the circumstances.[141]   The jury is charged that any "reasonable doubt," of whatever magnitude, must be resolved in favor of the accused. Such insistence on the greatest certainty that seems reasonably attainable can serve at the trial's end, like the presumption of innocence at the trial's start,[142] to affirm the dignity of the accused and to display respect for his rights as a person — in this instance, by declining to put those rights in deliberate jeopardy and by refusing to sacrifice him to the interests of others.

In contrast, for the jury to announce that it is prepared to convict the defendant in the face of an acknowledged and numerically measurable doubt as to his guilt is to tell the accused that those who judge him find it preferable to accept the resulting risk of his unjust conviction than to reduce that risk by demanding any further or more convincing proof of his guilt. I am far from persuaded that this represents the sort of thought process through which jurors do, or should, arrive at verdicts of guilt. Many jurors would no doubt describe themselves as being "completely sure," or at least as being "as sure as possible," before they vote to convict. That some mistaken verdicts are inevitably returned even by jurors who regard themselves as "certain" is of course true but is irrelevant; such unavoidable errors are in no sense *intended*,[143] and the fact that they must occur if trials are to be conducted at all need not undermine the effort, through the symbols of trial procedure, to express society's fundamental commitment to the protection of the defendant's rights as a person, as an end in himself. On the other hand, formulating an "acceptable" risk of error to which the trier is willing deliberately to subject the defendant would interfere seriously with this expressive role of the demand for certitude — however unattainable real certitude may be, and however clearly all may ultimately recognize its unattainability.

---

[141] See *The Supreme Court, 1969 Term*, 84 HARV. L. REV. 1, 157 & nn.8, 9 (1970).

[142] In some respects, it should be stressed, the insistence on certainty does *not* parallel the presumption of innocence. See note 137 *supra*.

[143] Tolerating a system in which perhaps one innocent man in a hundred is erroneously convicted despite each jury's attempt to make *as few mistakes as possible* is in this respect vastly different from instructing a jury to *aim* at a 1% rate (or even a .1% rate) of mistaken convictions. See Tribe 385–86, 388.

In short, to say that society recognizes the necessity of tolerating the erroneous "conviction of some innocent suspects in order to assure the confinement of a vastly larger number of guilty criminals" [144] is not at all to say that society does, or should, embrace a policy that juries, *conscious of the magnitude of their doubts in a particular case*, ought to convict in the face of this acknowledged and quantified uncertainty. It is to the complex difference between these two propositions that the concept of "guilt beyond a reasonable doubt" inevitably speaks. The concept signifies not any mathematical measure of the precise degree of certitude we require of juries in criminal cases,[145] but a subtle compromise between the knowledge, on the one hand, that we cannot realistically insist on acquittal whenever 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.[146]

   4. The Dehumanization of Justice. — Finally, we have been told by Finkelstein and Fairley that jurors using their method may find themselves "surprised" at the strength of the inference of guilt flowing from the combination of mathematical and non-mathematical evidence.[147] Indeed they may,[148] and in a far deeper sense than with other equally obscure forms of expert testimony, for such testimony typically represents no more than an input into the trial process, whereas the proposed use of Bayesian methods changes the character of the trial process itself. When that change yields a "surprisingly" strong inference of guilt in a particular case, it is by no means clear that, so long as one keeps one's numbers straight, "this . . . is no more than the evidence deserves." [149] Methods of proof that impose moral

---

[144] Dershowitz, *Preventive Detention: Social Threat*, TRIAL, Dec.-Jan. 1969–70, at 22.

[145] *Contra*, Finkelstein & Fairley 504; Ashford & Risinger, *Presumptions, Assumptions and Due Process in Criminal Cases: A Theoretical Overview*, 79 YALE L.J. 165, 183 (1969); Broun & Kelly, *supra* note 5, at 27.

[146] This seems to me a much more plausible account of the fuzziness of the "reasonable doubt" concept than does the alternative account that "courts shun responsibility for fixing a more precise threshold probability because they feel it should vary to some extent from case to case." Cullison, *supra* note 5, at 567. *See also* Broun & Kelly, *supra* note 5, at 31; Kaplan 1073.

[147] Finkelstein & Fairley 517; *see* p. 1358 *supra*.

[148] Suppose, for example, that each of three items of evidence, E_1, E_2, and E_3, has the effect of increasing a prior 1 percent suspicion of guilt ($P(X) = .01$) tenfold, so that $P(X|E_1) = P(X|E_2) = P(X|E_3) = .1$ If E_1, E_2, and E_3 are conditionally independent of X and not-X, *see* note 75 *supra*, then it turns out that $P(X|E_1 \& E_2 \& E_3)$ is *in excess of .93*, a result that might be counter-intuitive for many laymen. For another illustration, *see* RAIFFA, *supra* note 58, at 20–21.

[149] Finkelstein & Fairley 517.

---

blame or authorize official sanctions [150] on the basis of evidence
that fails to penetrate or convince the untutored contemporary
intuition threaten to make the legal system seem even more
alien and inhuman than it already does to distressingly many.
There is at stake not only the further weakening of the con-
fidence of the parties and of their willingness to abide by the
result, but also the further erosion of the public's sense that the
law's fact-finding apparatus is functioning in a somewhat com-
prehensible way, on the basis of evidence that speaks, at least in
general terms, to the larger community that the processes of
adjudication must ultimately serve. The need now is to enhance
community comprehension of the trial process, not to exacerbate
an already serious problem by shrouding the process in mathe-
matical obscurity.

It would be a terrible mistake to forget that a typical lawsuit,
whether civil or criminal, is only in part an objective search for
historical truth. It is also, and no less importantly, a ritual — a
complex pattern of gestures comprising what Henry Hart and
John McNaughton once called "society's last line of defense in
the indispensable effort to secure the peaceful settlement of so-
cial conflicts." [151]

One element, at least, of that ritual of conflict-settlement is
the presence and functioning of the jury — a cumbersome and
imperfect institution, to be sure, but an institution well calcu-
lated, at least potentially, to mediate between "the law" in the
abstract and the human needs of those affected by it. Guided
and perhaps intimidated by the seeming inexorability of numbers,
induced by the persuasive force of formulas and the precision of
decimal points to perceive themselves as performing a largely
mechanical and automatic role, few jurors — whether in criminal
cases or in civil — could be relied upon to recall, let alone to
perform, this humanizing function, to employ their intuition and
their sense of community values to shape their ultimate conclu-
sions.[152]

When one remembers these things, one must acknowledge
that there was a wisdom of sorts even in trial by battle — for at
least that mode of ascertaining truth and resolving conflict re-

---

[150] The argument I am here advancing applies with greatest force in the crim-
inal context, but it also has some significance in much ordinary civil litigation.

[151] Hart & McNaughton, *Evidence and Inference in the Law*, in EVIDENCE AND
INFERENCE 48, 52 (D. Lerner ed. 1958). I do not exclude the possibility that, in
extraordinary cases, and especially in cases involving highly technical controversies,
the "historical" function may be so dominant and the need for public comprehen-
sion so peripheral that a different analysis would be in order, laying greater stress
on trial accuracy and less on the elements of drama and ritual.

[152] *See* United States v. Spock, 416 F.2d 165, 182 (1st Cir. 1969).

flected well the deeply-felt beliefs of the times and places in which it was practiced.[153] This is something that can hardly be said of trial by mathematics today.

F. Conclusions

I am not yet prepared to say that the costs of mathematical precision enumerated here are so great as to outweigh any possible gain that might be derived from the carefully limited use of probabilistic proof in special circumstances. I do think it clear, however, that those circumstances would have to be extraordinary indeed for the proponents of mathematical methods of proof to make even a plausible case.

With the possible exception of using statistical data to shift the burden of production,[154] and perhaps with the further exception of using evidence as to frequencies in order to negate a misleading impression of uniqueness that expert opinion might otherwise convey,[155] I think it fair to say that the costs of attempting to integrate mathematics into the factfinding process of a legal trial outweigh the benefits. In particular, the technique proposed by Finkelstein and Fairley is incapable of achiev-

---

[153] *See generally* A. ENGELMANN, A HISTORY OF CONTINENTAL CIVIL PROCEDURE 155, 651–52 (1928).

[154] *See* p. 1361 & notes 33 and 102 *supra*.

[155] Some, but by no means all, of the costs of precision identified in this section are primarily costs for persons actually or potentially accused of crime. To the extent that a defendant in a criminal case wishes to employ mathematical methods in his own defense, these costs obviously weigh less heavily than they do in the case of prosecutorial use. One can imagine a variety of defensive uses of mathematics — for example, to establish the likelihood of an "accidental" cause of a seemingly incriminating event (as in the parking case, p. 1340 *supra*), or to show the likelihood that a person other than the accused committed the crime (as in the police and mistress cases, p. 1341 *supra*, modified by assuming different defendants from those there posited). But the most common defensive use would probably be the translation into quantitative form of an expert's damaging opinion that a certain physical trace or combination of traces must "almost certainly" have been left by the accused. *See* Finkelstein & Fairley 517. Courts otherwise hostile to probabilistic proof have at times allowed such quantification of expert opinion about trace evidence even at the prosecutor's initiation. *See* People v. Jordan, 45 Cal. 2d 697, 707, 290 P.2d 484, 490 (1955); *cf.* Miller v. State, 240 Ark. 340, 343–44, 399 S.W.2d 268, 270 (1966) (quantification rejected only because the prosecutor laid "no foundation upon which to base his probabilities"). Although the analysis of the preceding section would make me somewhat reluctant to accept such holdings (particularly in light of the "selection effect" described in note 40 *supra*), I am of the tentative view that the criminal defendant should nonetheless be permitted to initiate the quantification of this sort of expert opinion in order to establish a "reasonable doubt" as to his guilt. And, once such quantification has been initiated by the defense, the case for allowing the prosecution to rebut in mathematical terms becomes quite persuasive.

ing the objectives claimed for it, and possesses grave deficiencies that any other similarly conceived approach would be very likely to share.

It does not follow, however, that mathematical methods must play an equally limited role in the enterprise of designing trial procedures. What is true of mathematics as an aid to factfinding may be false of mathematics as an aid to rulemaking. In the pages that follow, I examine this separate issue and attempt to show that, although one can have somewhat more hope for mathematics in rulemaking, special problems of a quite serious character arise in that related context as well.

II. RULEMAKING WITH NUMBERS AND CURVES

A. One Simplified Model

Thus far, I have considered the role of mathematics in the process of proof, its potentialities and limitations in helping the trier of fact assess the probability of a disputed proposition. Once that probability has been assessed — by whatever means, mathematical or otherwise — there remains the problem of deciding what to do, what verdict to return. Can mathematical techniques be of assistance in formulating a rule of procedure — a standard of proof — that will solve that problem? More generally, what role can mathematics play in designing procedural rules for the trial process?

I want to consider first the narrower of those two questions, for it is the only one to which significant effort has thus far been directed. John Kaplan [156] and Alan Cullison [157] have both proposed a rather simple mathematical model of the trial process in order to determine the probability necessary to return a verdict. Although the model is as applicable to civil cases as to criminal, it is most readily understood in the setting of a criminal trial. They propose that a criminal trial be viewed as analogous to any other situation in which one must choose between two or more courses of action on the basis of a body of information which reduces, but does not wholly eliminate, the decisionmaker's uncertainty about the true state of the world and about the consequences in that world of any chosen strategy of conduct. [158]

In particular, the trier must choose between conviction (designated C) and acquittal (designated A) in the face of at least partial uncertainty as to whether the defendant is in fact guilty

---

[156] *See* Kaplan, *supra* note 5.
[157] *See* Cullison, *supra* note 5.
[158] *See generally* RAIFFA, *supra* note 58.

(designated G) or innocent (designated I). The four possible outcomes of the trier's decision problem are:

Outcome
(1) Convicting a guilty man, designated C_G
(2) Convicting an innocent man, designated C_I
(3) Acquitting a guilty man, designated A_G
(4) Acquitting an innocent man, designated A_I

The model posits that the rational trier should [159] choose C rather than A whenever the "expected utility" to the trier of the former choice would exceed the "expected utility" of the latter, in light of such factors as the seriousness of the offense, the severity of the punishment, and so on — much as a rational gambler would select the bet that maximizes his expected gains, taking into account his present position, his needs, and his attitudes toward risk.[160]

In order to make the necessary choice, the trier must first decide how much he would like or dislike each of the four possible outcomes of the proceeding — that is, he must decide what "utility" each has for him.[161] Suppose that the trier's order of preference, from the outcome he would like best to the one he would like least, is C_G, A_I, A_G, C_I. In order to assign quantitative utilities $U(C_G)$, $U(A_I)$, $U(A_G)$, and $U(C_I)$ to these outcomes, he begins by assigning a maximum utility of 1 to the outcome he likes most and a minimum utility of 0 to the outcome he likes least:

$$U(C_G) = 1$$
$$U(C_I) = 0$$

To decide what utility between 0 and 1 to assign A_I, the trier

---

[159] It is not entirely clear to what extent the model is intended by Kaplan or Cullison as a description of how trial decisions are in fact made, *see, e.g.,* Kaplan 1069–70, 1075, to what extent it is offered as an heuristic device for illuminating the trial process, *see, e.g., id.* 1066, 1091, and to what extent it is meant as a normative guide to the trier's choice of verdict, *see, e.g., id.* 1065, 1072–74, 1092. I am concerned here with the model's heuristic and normative roles only, and of the four criticisms I later advance, *see* pp. 1381–85 *infra,* all but one, *see* p. 1384 *infra,* apply to the former as well as to the latter.

[160] It might be objected that the gambling analogy is a weak one insofar as the payoffs in the trial "game" accrue directly to persons other than the decision-maker. Since there are obviously significant psychological payoffs for the gambler as well, however, the objection seems to me a superficial one. *Cf.* note 60 *supra.*

[161] An alternative approach is possible, focusing simply upon the "disutility" of each of the two possible kinds of errors (erroneous conviction or erroneous acquittal), *see* note 168 *infra,* but expressing the problem in the terms used here facilitates the assignment of numbers to the various outcomes and makes somewhat more transparent the "cognitive dissonance" problem *discussed at* pp. 1383–84 *infra.*

asks himself such questions as the following: Would I rather get
A_I for sure or get a $1/2$ chance of the best outcome, C_G, and a
$1/2$ chance of the worst, C_I? If the answer is that he would
rather get A_I, then $U(A_I)$ is said to exceed $1/2$; if he would
prefer the gamble, $U(A_I)$ is less than $1/2$. If it turns out that
$U(A_I)$ exceeds $1/2$ by this test, then the trier asks himself
whether he would rather get A_I for sure or get, say, a $3/4$ chance
of C_G and a $1/4$ chance of C_I. If the answer this time is that
he would rather take the chance, then $U(A_I)$ falls between $1/2$
and $3/4$. In this way, the trier "closes in" on $U(A_I)$ until he
ultimately pinpoints its value. To say, for example, that $U(A_I)$
$= 2/3$ is to say that the trier would be as satisfied getting A_I
for sure as he would be getting a $2/3$ chance of C_G and $1/3$
chance of C_I. Suppose for the sake of illustration that, by this
same process, the trier concludes that, for him, $U(A_G) = 1/2$.

Now the trier is in a position to decide how sure of the de-
fendant's guilt he would have to be before preferring C to A.
To that end, let P designate the trier's probability assessment
of G in light of all the evidence in the case. Then, if the trier
chooses C, there is a probability of P that he will get C_G and a
probability of $1{-}P$ that he will get C_I. $EU(C)$, the "expected
utility" or "expected desirability" of this choice, is the sum of
two products: (1) the probability of guilt, P, multiplied by the
desirability of C_G; and (2) the probability of innocence, $1{-}P$,
multiplied by the desirability of C_I:

$$EU(C) = P \cdot U(C_G) + (1{-}P) \cdot U(C_I) .$$

But this simply equals P, since $U(C_G) = 1$ and $U(C_I) = 0$. If
the trier chooses A, there is a probability of P that he will get
A_G and a probability of $1{-}P$ that he will get A_I. $EU(A)$, the
"expected utility" of A, is thus

$$EU(A) = P \cdot U(A_G) + (1{-}P) \cdot U(A_I) ,$$

which in our case equals $P \cdot 1/2 + (1{-}P) \cdot 2/3$, or $\dfrac{4{-}P}{6}$. Thus

the expected utility of choosing C exceeds that of choosing A
whenever P exceeds $\dfrac{4{-}P}{6}$, which occurs whenever P exceeds $4/7$.

Hence, given the utilities the trier has assigned to the four pos-
sible outcomes, the model supplies him with a rule of procedure
for this criminal case: "Consider the evidence and then vote to
convict if and only if you think that the probability of the de-
fendant's guilt exceeds $4/7$." [162]

---

[162] It should be noted that the procedural rule produced by· the model will

Now, one might have qualms about the resulting procedural rule — because one regards a threshold probability of 4/7 as much too low, or because one objects in principle to the willing taking of *any* measurable risk of convicting an innocent man,[163] or because one regards as unacceptable the cost of openly *announcing* that willingness [164] — but my concern here is not so much with the result as with the method used to arrive at it. It is to a criticism of that method that the next section is addressed.

B. A Critique of the Model

1. *Misspecification of Consequences.* — The model described above assumes the existence of meaningful answers to such questions as: "How much would you regret the erroneous conviction of this defendant for armed robbery?" But the answer must surely be "It depends." It depends in part upon the *character* of the error itself; mistaken identity might be worse, for instance, than misjudged intention and worse still than a miscalculated statute of limitations.[165] And it depends even more significantly upon the *process* that led to the error; one cannot equate the lynching of an innocent man with his mistaken conviction after a fair trial. Indeed, it is at least arguable that there is nothing good or bad about *any* trial outcome as such; that the *process*, and not the *result in any particular case*, is all-important. To be sure, some concern for the mix of correct and erroneous outcomes operates as a constraint on what might other-

---

always have this numerical form; it can never assume the indefinite shape of "subjective certitude" or "guilt beyond a reasonable doubt." *Cf.* pp. 1374–75 *supra*. In his concurring opinion in *In re* Winship, 397 U.S. 358, 368 (1970), Mr. Justice Harlan argued that the "reasonable doubt" standard in criminal cases, like the quite different "preponderance of the evidence" standard in much civil litigation, merely reflects an assessment of the comparative social disutility of erroneous acquittal and erroneous conviction. *Id.* 370–71. Given the societal recognition that the latter error is far worse than the former, *id.* 372, a demanding burden of proof is imposed on the prosecution in order to assure that men are wrongly convicted much less often than they are wrongly acquitted. *Id.* at 371. *See also* Ball, *supra* note 37, at 816. This analysis, for which Mr. Justice Harlan credits Kaplan, *supra* note 5, 397 U.S. at 370 n.2, suffers from all of the defects I will shortly discuss with respect to Kaplan, *see* pp. 1381–85 *infra*, and suffers in addition from the defect that it proves too little. Specifically, the objective of assuring that erroneous acquittals of the guilty occur with greater frequency than erroneous convictions of the innocent demands only that the prosecution be required to prove its case more convincingly than must a civil plaintiff (*e.g.*, by "clear and convincing evidence," or perhaps "to a probability of 9/10"), *not* that it produce the "subjective state of certitude" stressed by the Court's majority opinion. *See The Supreme Court, 1969 Term*, 84 HARV. L. REV. 1, 158 n.13 (1970).

[163] *See* pp. 1372–74 *supra*.

[164] *See* note 140 *supra*.

[165] *See* Kaplan 1073.

wise be deemed acceptable trial procedures — but the accepta-
bility of a process is not simply a function of the number of
correct or erroneous convictions or acquittals it yields. At the
very least it is clear that our preferences, and those of the trier,
attach not to the bare consequences of correct or erroneous
conviction or acquittal. They attach instead, and properly so,
to the consequences — for a broad range of values and interests
— of the defendant's correct or erroneous conviction or acquittal
after a given sort of trial, operating with a particular set of rules
and biases, and governed by a specific standard of proof.[166]

In particular, the trier might justly regard as worse the
erroneous conviction of a man to whose guilt he had attached
a probability of just over $4/7$ than the erroneous conviction
of one whose guilt had seemed to be virtually certain. Indeed,
the trier would probably have to attach a different "utility"
to the outcome of erroneously convicting a man on the basis of
a standard that appeared to convey to the community at large a
willingness to take calculated risks of such errors, than he would
to the outcome of erroneously convicting a man on the basis of
a standard that gave no such appearance.[167] At a minimum,
therefore, because the utilities of the various consequences would
themselves be functions of the apparent probability of the de-

---

[166] To illustrate the sharp difference between this view and the model put forth
by Kaplan, note how Kaplan explains why our legal system typically excludes
evidence of previous convictions from the prosecution's case-in-chief. This is done,
he says, because including such evidence might lead the jurors "to the perhaps
rational but clearly undesirable conclusion that because of his earlier convictions,
D_1, the disutility of convicting the defendant should he be innocent, is minimal,"
Kaplan 1074, and that a low probability of present guilt should thus suffice to
warrant his conviction. *Id.* 1077. But if it were simply a matter of fitting the
standard of proof to the comparative utilities and disutilities of the four possible
outcomes, *why* would that conclusion be "clearly undesirable"? Because, we are
told, ours is "a system of justice that regards it as crucial that the defendant be
found guilty only of the crime specifically charged." *Id.* 1074. Yet, if that is so,
and if a conclusion that flies in its face nonetheless emerges as "perhaps rational"
and indeed inevitable within the four corners of Kaplan's utilities, must one not
conclude that the model built on those four utilities is inherently defective? Our
system typically excludes prior convictions (with, of course, many exceptions) for
the kinds of reasons that any adequate model of the criminal trial must somehow
reflect — for reasons of repose; for the prevention of multiple punishment; for the
appearance of fairness; for the preservation of a substantive system of law in
which the accused, however long his record, can by his own choice avoid future
entanglement with the criminal process, *cf.* Tribe 394–96; and for the preservation
of a procedural system of law in which the accused, whatever his background, is
given a well-defined opportunity to rebut a precise charge. *Cf. id.* 392–94. All
of those factors enter into the question whether a man's trial was a "fair" one;
none of them figures in the simplistic calculation of how desirable or undesirable
would be each of its four possible outcomes.

[167] *See* note 140 *supra.*

fendant's guilt, any equation designed to compute the threshold probability above which conviction would be preferable to acquittal would have to be far more complex than Kaplan and Cullison have supposed.[168] And that, in turn, could preclude the existence of *any* single threshold and would in any event make the model, already too obscure for actual use by a trier of fact, more esoteric still.

2. *Problems of Cognitive Dissonance.* — The preceding discussion demonstrates that a legal trial differs from the usual sort of management problem to which utility theory has previously been applied[169] in at least one important respect: various features of the procedure followed to reach the decision, including the standard of proof applied, are themselves integral parts of the consequences to be optimized, a fact that greatly complicates the optimization process. The trial decision also differs from the classical management problem in another crucial respect: the decisionmaker will invariably have preferences not only with respect to the consequences of his choice but also with respect to the underlying facts themselves, facts over which he can exercise no control. Thus, for example, the trier's reluctance to see an innocent man put to the ordeal of trial and his wish to avoid discovering that the man who is in fact guilty remains at liberty may combine to reduce the desirability, for him, of the outcome previously designated A_I (acquittal of the defendant who is in fact innocent), as compared with the outcome previously designed C_G (conviction of the defendant who is in fact

---

[168] The Kaplan-Cullison model, to generalize the computation performed at pp. 1379-81 *supra*, yields the rule that conviction should be preferred to acquittal whenever P, the final probability estimate of the defendant's guilt, exceeds the quotient:

$$\dfrac{1}{1 + \dfrac{U(C_G) - U(A_G)}{U(A_I) - U(C_I)}}$$

Kaplan designates the difference $U(C_G) - U(A_G)$ by the symbol D_g and the difference $U(A_I) - U(C_I)$ by the symbol D_i. *See* note 161 *supra*. If, as I have suggested in text, the values of $U(C_G)$, $U(A_G)$, $U(A_I)$, and $U(C_I)$ themselves depend on P, then the rule will have a more complex form. Thus, if the utilities or at least their differences $(U(C_G) - U(A_G)$ and $U(A_I) - U(C_I))$ depend in a linear way on P, there will exist numbers U_1, U_2, and U_3 such that conviction is preferable to acquittal whenever $U_1 P^2 + U_2 P + U_3 > 0$. Because such an equation can have two roots between 0 and 1, there may be no single threshold value P^* such that conviction is preferable to acquittal for all $P \geq P^*$.

[169] A typical example would be the question whether or not to drill for oil at a given site before one's option expires, given incomplete information about such variables as the cost of drilling and the extent of oil deposits at the site. *See* RAIFFA, *supra* note 58, at xx.

guilty).[170] The risk is that the trier will not only allow his hope that the accused is in fact guilty to influence his perception of the evidence — a classic case of adjusting cognition to avoid psychological dissonance [171] — but will also allow that hope, through a distortion in the comparative magnitudes of $U(A_I)$ and $U(C_G)$, to influence his determination of what standard of proof to apply. The suggested method for arriving at that standard in no way guards against this danger.

   3. Positing the Wrong Decisionmaker. — It may be that *no* method of arriving at the standard of proof can avoid the problem identified above unless it effectively separates the selection of that standard from the decision of the particular case. Indeed, a variety of other important values, including that of both real and apparent equality in the treatment of accused persons, all point in the same direction: the factfinder in a criminal trial should not be encouraged to do what the proposed model demands of him — namely, that he expressly assess the desirability or "utility" of the various consequences that might flow from correctly or erroneously convicting or acquitting the particular defendant then on trial.[172]

   The fact that such matters as the defendant's reputation and the likely sentence, all of which, of course, bear directly on the utilities of the four possible outcomes, are nonetheless often kept from the jury "is hard to defend . . . on a decision-theoretic view" [173] only if one's decision theory neglects to ask about the institutional competencies of the several elements of the legal system. But when one gives due weight to the costs of combining in the trier the separate functions of deciding what happened in a particular case and evaluating the anticipated consequences of alternative verdicts, one will expect the lawmaker rather than the factfinder to use a model such as the one Kaplan and Cullison propose, and one will define the decision problem to be

---

   [170] An analogous problem outside the trial context would be presented by a choice among alternative medical strategies, all entailing some risks, for a patient who might or might not have cancer. The application of classical techniques of decision-analysis to such situations, in which the decisionmaker cannot be entirely neutral with respect to the uncertain facts underlying his problem, is a matter of much current interest and research within the decision-analysis profession, although I am aware of no published discussions of the problem thus far.

   [171] *See generally* L. Festinger, A Theory of Cognitive Dissonance (1957).

   [172] *But see* note 177 *infra.* I would not have as much quarrel with techniques that called upon the factfinder to think in less formal terms about how high a probability of guilt to require as a precondition of returning a verdict against the defendant, but any method yielding a numerical conclusion to that question would be subject to the basic objection I have made to attempts at quantifying the final probability of guilt. *See* pp. 1372–75 *supra.*

   [173] Kaplan 1075.

solved not as the one-shot problem of fixing a standard of proof for a particular trial with four possible outcomes, but as the much larger problem of establishing such standards for the trial system as a whole.

4. Operating in a Factual Vacuum. — Having thus broadened the inquiry, one cannot avoid noticing that the model proposed by Kaplan and Cullison is, oddly enough, structured to make use of *none* of the crucial facts one would surely want to know when establishing a standard of proof. We are, after all, talking about some real things: crimes, their prevention, the incapacitation of their perpetrators, and the protection of innocent persons from being falsely convicted for them. If it be proposed that juries in a certain kind of case should convict whenever they think a defendant's probability of guilt exceeds 4/7, no one concerned with these real things could fail to ask questions such as these: How many guilty men are likely to be erroneously acquitted under that standard? How many innocent men are likely to be erroneously convicted? What will be the effect on the likely number of offenses? What will be the impact on the fear of false prosecution and unjust imprisonment? The answers to such questions, in turn, will depend on such other inquiries as these: How easy or difficult is it for the state to make the probability of an innocent man's guilt appear to exceed 4/7? How easy is it for a guilty defendant to make the probability of his innocence appear to exceed 3/7? How many innocent men are brought to trial? How might the number of innocents tried depend on the announced standard of proof? How do the number of offenses or the fear of erroneous conviction relate to the probability of conviction if guilty? To the probability of conviction if innocent? To the ratio of convictions to acquittals? To the absolute number of convictions? To the absolute number of acquittals?

The striking thing is that the answers to virtually none of these obviously relevant questions could ever find their way into the Kaplan-Cullison model, for the answers simply do not relate, in the main, to an assessment of the desirability or undesirability of one or another outcome of a paradigm trial; they relate instead to the characteristics of a much broader system.[174] The proposed model, then, is not really useful and does not provide

---

[174] The answer to such an obviously crucial question as "how much deterrent effect will flow from convicting whenever the probability of guilt exceeds 4/7" cannot influence the trier's assessment of the desirability or undesirability of correctly or erroneously convicting or acquitting any given defendant, and hence cannot affect the decision in the Kaplan-Cullison model whether to treat a 4/7 probability of guilt as sufficient to convict.

a fair test for the potentialities of mathematical methods in procedural design.

C. More Sophisticated Techniques

A somewhat fairer test might be provided by an approach employing what economists usually call "choice sets" and either "indifference curves" or "preference contours." [175] For a particular crime, the rulemaker would first establish the "choice set" open to him by investigating the functional relationship to be expected between the percentage of guilty convicted and the percentage of innocent convicted, recognizing that — at any given level of resource investment — convicting more of the guilty may require relaxing various procedural standards (including, but not limited to, the probability of guilt required for conviction in any particular case) and thereby convicting more of the innocent as well. That functional relationship will reflect, among other things, the ratio of guilty to innocent defendants among those brought to trial for the crime in question, the sensitivity of trial outcomes to various procedural rules, and a number of other factors that would surely vary from one jurisdiction to another, and from one crime to another.

Having thus established in some empirical way the choices open to him with respect to this crime, the rulemaker would next think about his preferences, or those of his constituents. If he could convict, for example, 60% of the guilty with a 1% chance of convicting an innocent, how much would he let the latter percentage rise in order to convict an additional 5% of the guilty? To convict an additional 10% of the guilty? In thinking about the answers to such questions, the rulemaker would, of course, have to take account of whatever information he could develop on such topics as the relationship between the probability of conviction if guilty and the corresponding frequency of offenses,[176] and he would also have to take account of such factors as the relationship between the probability of conviction if inno-

---

[175] I am much indebted to Professor Thomas C. Schelling for helping me translate my earlier and more intuitive formulation of this general approach into the one employed in the present article. Techniques of a related sort are employed by both Becker, *supra* note 5, and Birmingham, *supra* note 5.

[176] Of course it might turn out that the frequency of offenses depends strongly not only on the probability of conviction if guilty but also on the ratio of convictions to trials, or even on the absolute number of convictions. If this should prove to be the case, a rulemaker using this sort of approach might be tempted to design the system so as to convict more innocents, not only as an unavoidable cost of convicting a higher percentage of guilty, but also as part of a deliberate strategy of deterrence. Needless to say, this may well be a decisive objection to this form of analysis.

cent and the corresponding level of fear and insecurity among his constituents.[177]

Starting with any arbitrarily chosen point — such as the one at which 60% of the guilty and 1% of the innocent are convicted — the rulemaker would consider such relationships in deriving his "preference contour" through the point (60, 1):
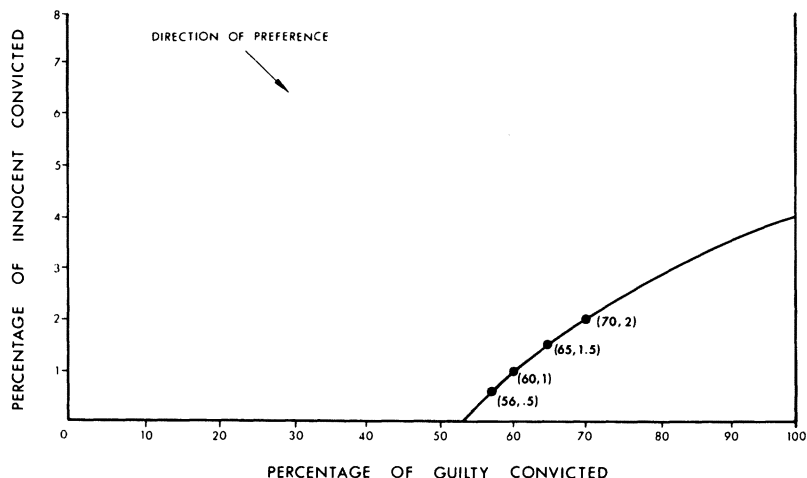

The curve drawn indicates, among other things, that the rulemaker would be willing to let the figure of 1 climb to 1.5 (but no higher) in order to increase the 60 to 65, and to 2 (but no higher) in order to increase the 65 to 70. It also indicates that he would be willing to let the 60 drop to 56 (but no lower) in order to decrease the 1 to .5.

The rulemaker would then draw another preference contour starting at the point (60, 2); another one starting at (60, 3);

---

[177] If, as one might well suspect, the constituents who matter most to the rulemaker, *see* V.O. KEY, AMERICAN STATE POLITICS: AN INTRODUCTION 140–41 (1956), will themselves be insulated by social or economic status from the "insecurity costs" of a rising risk of convicting innocents, the rules arrived at through the procedure outlined here (and perhaps, though by no means certainly, the rules that would be arrived at in less calculating ways as well) will expose categories of persons more susceptible to false arrest and mistaken conviction to a greater danger of such misfortunes than would result from rules chosen by, or on behalf of, men acting under a veil of ignorance as to their ultimate status in the society they are designing. *See* Rawls, *Justice as Fairness*, 67 PHILOS. REV. 145 (1958). The jury, on the other hand, is — or can more readily be made into — a body of persons more likely than the typical legislator's important constituents to pay in false convictions the price of relaxed prosecutorial burdens. This may create a powerful reason for leaving to the jury broader discretion with respect to such matters as standards of proof than was argued for at p. 1384 *supra*, at least so long as the delegation of such discretion takes a sufficiently inarticulate form.

and so on, thereby building up a complete set of indifference
curves or preference contours:[178]

PERCENTAGE OF GUILTY CONVICTED
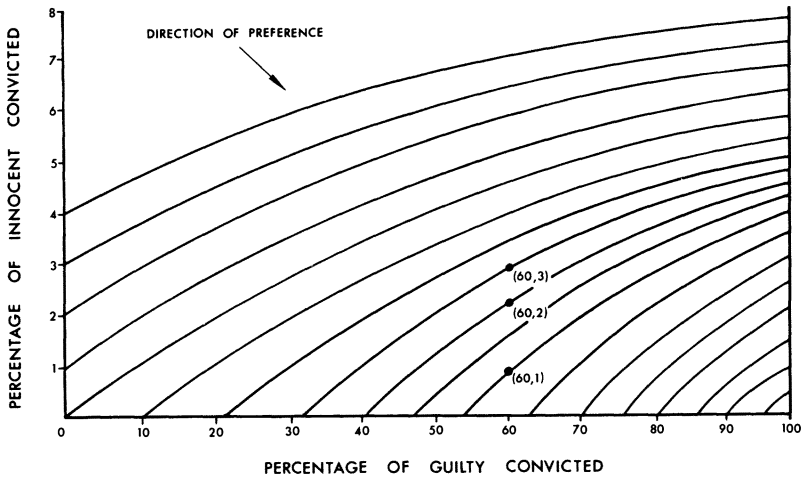
Finally, the rulemaker would superimpose on these prefer-
ence contours the empirically established "choice set" C — the
functional relationship telling him, at the assumed level of re-
source investment, how many innocents would be convicted for
any given percentage of guilty:

PERCENTAGE OF GUILTY CONVICTED
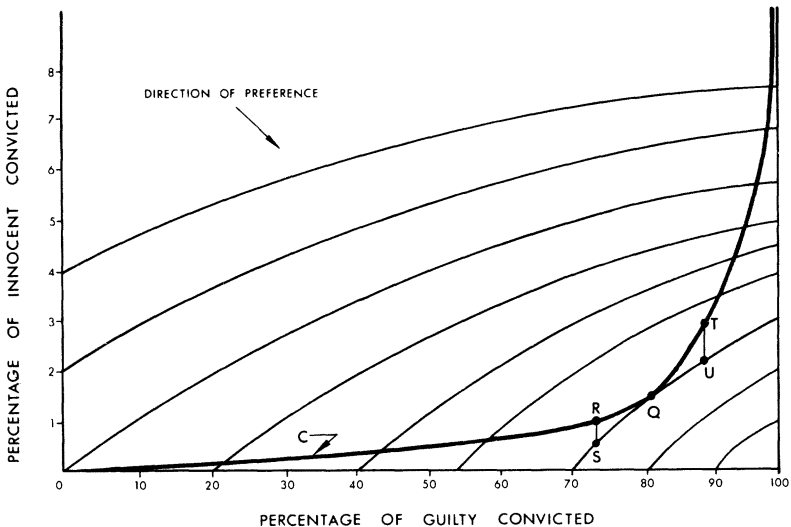
The optimum point on the choice set is Q, the point of tan-
gency between that set and a preference contour.[179] By finding

---

[178] Each rulemaker has, of course, an infinite number of these preference con-
tours for a given crime; in the above illustration, only a representative subset of
the entire family of contours can be depicted.

[179] To see why Q is optimal, simply imagine any alternative points on the

that point of tangency, the rulemaker determines the percentage of guilty convictions he should aim for and knows the
corresponding percentage of innocent convictions that will result. If, for example, Q is at the point (80, 1.2), the rulemaker
knows that he should design the procedure for trials of the
crime in question so as to convict some 80% of the guilty at a
cost of convicting some 1.2% of the innocent. The problem of
discovering just what combination of procedures (standards of
proof, presumptions, rules of admissibility and exclusion, and the
like) will have that approximate effect then becomes the next
task — obviously a difficult one — on the rulemaker's agenda
for empirical research and mathematical analysis.

D. Some Tentative Reservations

1. On Precision and Quantification. — To whatever extent
all of this represents, albeit in somewhat simplified and preliminary form, a typical instance of mathematical reasoning in the
design of trial procedures, it is important to explore the costs
that may be incurred by its use.

Those costs, in large measure, are the same costs of precision
that I have examined in another context.[180] In particular, there
is a significant risk that the greater ease with which the rulemaker will be able to quantify some variables (such as the incidence of crime) as compared to others (such as the insecurity flowing from fear of false conviction) will skew his decision in unfortunate directions, leaving serious doubt whether the exactitude of
the numbers and curves will, in the end, lead to better rules.[181] In
the preference contour exercise attempted above, for example, it
seems clear that the shape of the contours ought at least to reflect the procedures and the philosophy that would be required
to achieve any given mix of trial outcomes. One cannot really
say, for example, whether one feels better or worse about con

---

choice set C such as R or T. Because S, Q, and U in the above illustration lie
on the same preference contour or indifference curve, the rulemaker feels equally
satisfied at those three points. But he clearly feels better at S than at R and
better at U than at T, since in each case the first point of the pair provides the
same benefit as the second (the same percentage of guilty convicted) at lower cost
than the second (a lower percentage of innocent convicted), assuming that the
perversity described in note 176 *supra*, does not obtain. Since the rulemaker is
indifferent as among S, Q, and U, and prefers S to R and U to T, he must prefer
Q to either R or T, which implies that Q is indeed optimal. It should be noted,
however, that the ability of this model to yield a unique optimum depends upon
several assumptions with respect to the shape of the preference contours, that of
the choice set, and the relationships among them.

[180] *See* pp. 1358–77 *supra.*
[181] *Cf.* pp. 1361–66 *supra.*

victing 80% of the guilty and 1.2% of the innocent than one feels about convicting 90% of the guilty and 2.5% of the inno- cent, unless one knows how trial procedures might have to be altered [182] in order to go from the former point to the latter. But the costs of that procedural alteration, in terms of the many intangible consequences of such a change for a broad spectrum of values, will almost certainly prove harder to quantify than will the benefits of convicting another 10% of the guilty.[183] As a result, the preference contours may fail to reflect how the rule- maker really *does* feel about things — and the conclusion to which they point may be less acceptable than one more intuitively and impressionistically derived.

Moreover, once one is precise and calculating about rule- making, one can no longer so easily enjoy the benefits of those profoundly useful notions — like the "presumption of innocence" and "acquittal in all cases of doubt" — that we earlier saw threatened by mathematical proof.[184] After deciding in a de- liberate and calculated way that it is willing to convict twelve innocent defendants out of 1000 in order to convict 800 who are guilty — because that is thought to be preferable to convicting just six who are innocent but only 500 who are guilty — a com- munity would be hard pressed to insist in its culture and rhetoric that the rights of innocent persons must not be deliberately sacri- ficed for social gain.[185]

There are, finally, several problems of a different order — problems that go to the wisdom of being somewhat fuzzy and open-ended in one's statement of at least some kinds of standards and procedures that are designed to guide others over time. I have in mind the great advantage in some areas of principles over rules,[186] of formulations that facilitate consensus on re-

---

[182] As, for example, by relaxing the privilege against self-incrimination, or the requirement of proof beyond a reasonable doubt.

[183] Particularly is this so in light of the expressive role of procedure discussed at pp. 1391–93 *infra*.

[184] *See* pp. 1370–75 *supra*.

[185] Although I regard this as an important problem, I do not think it quite as significant as the analogous problem in the context of mathematical proof, where the decision to take a visible and calculated risk of erroneously convicting a specific accused person is more dramatic, may be thought to entail a lack of respect for the accused as an individual person, and seems more likely to have wide-ranging psychological impact. *See* pp. 1372–75 *supra*. *Cf.* Tribe, *supra* note 133, at 386 n.65.

[186] *See, e.g.*, Dworkin, *The Model of Rules*, 35 U. CHI. L. REV. 14 (1967). *See also* p. 1375 *supra* and note 162 *supra*. This is not to deny, of course, that "bright line" rules are occasionally preferable, *see, e.g.*, Bok, *Section 7 of the Clayton Act and the Merging of Law and Economics*, 74 HARV. L. REV. 226, 270– 73, 350–55 (1960), but only to stress the importance of being able to choose general principles when they seem better suited to one's purposes.

sults [187] and leave one free to move in many different directions as one's understanding grows and as one's needs evolve.[188] There is no necessary reason why mathematical analysis, operating with deliberately unspecified variables, cannot someday prove helpful in this subtle business — but I doubt that the day has come. At least in the rudimentary state of the art represented by the preceding two sections, the mission of mathematics *is* specification, and the almost inevitable corollary of its serious use in these circumstances is a move away from the open-ended to the rigorously defined.

2. On Utility and Ritual. — The appropriateness of applying mathematical methods to decisionmaking seems clearest when the alternative acts among which one is deciding are significant only as means to some external set of agreed-upon ends. For the decisionmaker can then approach his problem as the essentially mechanical one of choosing the act whose expected consequences will maximize a suitably weighted combination of those ends, subject to some appropriately defined set of constraints.

The great difficulty with thinking in this way about the choice of legal rules and the design of legal institutions is that such rules and institutions are often significant, not only as means of achieving various ends external to themselves, but also as ends in their own right, or at least as symbolic expressions of certain ends and values.

As much of the preceding analysis has indicated,[189] rules of trial procedure in particular have importance largely as expressive entities and only in part as means of influencing independently significant conduct and outcomes.[190] Some of those rules, to be sure, reflect only "an arid ritual of meaningless form," [191] but others express profoundly significant moral relationships and principles — principles too subtle to be translated into anything less complex than the intricate symbolism of the trial process. Far from being either barren or obsolete, much of what goes on in the trial of a lawsuit — particularly in a criminal case — is partly ceremonial or ritualistic in this deeply positive sense, and partly educational as well; procedure can serve a vital role as conventionalized communication among a trial's participants, and as something like a reminder to the community of the principles

---

[187] *See generally* C. LINDBLOM, THE INTELLIGENCE OF DEMOCRACY 207–08 (1965) (a study of decisionmaking through mutual adjustment).

[188] *See, e.g.,* Freund, *Privacy: One Right or Many,* 13 NOMOS 182 (1971).

[189] *See* pp. 1370–71 & 1372–76 *supra.*

[190] *See, e.g.,* C. FRIED, AN ANATOMY OF VALUES 125–32 (1970); E. GOFFMAN, INTERACTION RITUAL 10–11, 19, 54 (1967). *See also* J. Feinberg, *The Expressive Function of Punishment,* 49 THE MONIST 397 (1965), in DOING AND DESERVING (1970). For much of the discussion that follows, I am heavily indebted to the work of both Professor Goffman and Professor Fried.

[191] Staub v. City of Baxley, 355 U.S. 313, 320 (1958).

it holds important.[192] The presumption of innocence,[193] the rights to counsel [194] and confrontation,[195] the privilege against self-incrimination,[196] and a variety of other trial rights,[197] matter not only as devices for achieving or avoiding certain kinds of trial outcomes,[198] but also as affirmations of respect for the accused as a human being — affirmations that remind him and the public about the sort of society we want to become and, indeed, about the sort of society we are.[199]

Perhaps these expressive roles of procedure can be formally assimilated into a utility-maximizing model by adding on appropriate values to the weighted combination of preferred ends.[200] But, however completely this amplification of the model mirrors all of one's values, there is little chance of capturing the fact that much of what matters about expressive rules, procedural or otherwise, is that they *embody* and do not merely *implement* the values of the community that follows them. To employ mathematical techniques to help choose that rule which will maximize an appropriately weighted mix of certain values or preferences is to take those values as given — as objects outside the rules among which one is choosing. In fact, however, the very *choice* of one rule rather than another — of a rule that the accused cannot be forced to testify against himself, for example — may itself evidence and indeed constitute a *change* in the mix of basic values of the society that has made the choice in question.[201] At this point, the decision problem — if it can

---

[192] As Thurman Arnold once observed,

> Trials are like the miracle or morality plays of ancient times. They dramatically present the conflicting moral values of a community in a way that could not be done by logical formalization. Civil trials perform this function as well as do criminal trials, but the more important emotional impact upon a society occurs in a criminal trial.

Arnold, *The Criminal Trial as a Symbol of Public Morality*, in CRIMINAL JUSTICE IN OUR TIME 141–42, 143–44 (A. Howard ed. 1965).

[193] *See* pp. 1370–71 *supra*.

[194] *See, e.g.*, Gideon v. Wainwright, 372 U.S. 335 (1963).

[195] *See, e.g.*, Pointer v. Texas, 380 U.S. 400 (1965).

[196] *See, e.g.*, Malloy v. Hogan, 378 U.S. 1 (1964).

[197] *E.g.*, the defendant's right in some circumstances to exclude evidence of prior crimes, discussed in note 166 *supra*.

[198] *E.g.*, fewer erroneous convictions.

[199] *See, e.g.*, pp. 1370–71 & 1373–75 *supra*.

[200] *See, e.g.*, Nozick, *Moral Complications and Moral Structures*, 13 NAT. L.F. 1, 3 (1968), *discussed in* C. FRIED, AN ANATOMY OF VALUES 95, 157 (1970).

[201] For example, the illustration given by Nozick, *id.*, proposes that for any theory T that describes which actions are morally impermissible, one may define a function f whose maximization mirrors the structure of T by setting $f(A) = 0$ whenever action A is impermissible according to T and $f(A) = 1$ otherwise. But if some A's have the perverse effect of *changing* T, then even this "gimmicked-up" real-valued function will not do as a function whose maximization mirrors the moral values implicit in T. The character of at least some procedural rules, I am

still be called that — is to "choose" what fundamental values one wants to have and not simply to find the best way of implementing a set of values accepted as given.[202] Numbers and curves can be of relatively little use at so ultimate a level.

E. Conclusions

Reluctant as I am to make confident pronouncements about the final limits of mathematics in the fact-finding process of a civil or criminal trial,[203] I am more reluctant still to attempt any definitive assessment of how far mathematical methods and models can acceptably be exploited in the rulemaking process that determines how trials are conducted.

I have examined in some detail one simple model proposed by Kaplan and Cullison to assist in the determination of standards of proof, and have concluded that their approach, like that of Finkelstein and Fairley in the context of mathematical evidence, is more misleading than helpful. I have analyzed less closely the outlines of a more complex methodology — one that would apply preference contours and choice sets to the derivation of rules of criminal procedure — and have found that methodology substantially more enlightening but still far from satisfactory. And I have attempted to show, finally, that there may be at least some inherent limitations in the linking of mathematics to procedural rulemaking — limitations arising in part from the tendency of more readily quantifiable variables to dwarf those that are harder to measure, in part from the uneasy partnership of mathematical precision and certain important values, in part from the possible incompatibility of mathematics with open-ended and deliberately ill-defined formulations, and in part from the intrinsic difficulty of applying techniques of maximization to the rich fabric of ritual and to the selection of ends as opposed to the specification of means.

In an era when the power but not the wisdom of science is increasingly taken for granted, there has been a rapidly growing interest in the conjunction of mathematics and the trial process. The literature of legal praise for the progeny of such a wedding has been little short of lyrical. Surely the time has come for someone to suggest that the union would be more dangerous than fruitful.

---

suggesting, is related to our value system precisely as T-changing A's are related to T.

[202] Even if one takes the view that means and ends (or values) differ not in kind but only in degree, this argument still has significance as indicative of how extraordinarily little can be taken as "given," and hence as subject to weighted maximization, in the procedural area.

[203] *See* p. 1377 *supra*.