Veridical Verdicts: Increasing Verdict Accuracy Through the Use of Overtly Probabilistic Evidence and Methods

Jonathan J. Koehler

Daniel N. Shaviro

VERIDICAL VERDICTS: INCREASING VERDICT ACCURACY THROUGH THE USE OF OVERTLY PROBABILISTIC EVIDENCE AND METHODS

Jonathan J. Koehler †
Daniel N. Shaviro ††

For nearly twenty years, law journals have been the forum for a bitter debate about the use at trial of overtly probabilistic evidence and methods.[1] The debate embraces (and often confuses) at least three issues.

The first issue is the appropriateness of allowing juries to base factual determinations, in whole or in part, on probabilities derived from base rate evidence.[2] While some commentators generally support the use of base rate evidence at trial,[3] others argue that it is

---

1 *See, e.g.,* L. JONATHAN COHEN, THE PROBABLE AND THE PROVABLE (1977) [hereinafter L.J. COHEN 1]; V.C. Ball, *The Moment of Truth: Probability Theory and Standards of Proof*, 14 VAND. L. REV. 807 (1961); Lea Brilmayer, *Second Order Evidence and Bayesian Logic*, 66 B.U.L. REV. 673 (1986); Lea Brilmayer & Lewis Kornhauser, *Review: Quantitative Methods and Legal Decisions*, 46 U. CHI. L. REV. 116 (1978); James Brook, *Inevitable Errors: The Preponderance of the Evidence Standard in Civil Litigation*, 18 TULSA L.J. 79 (1982) [hereinafter "Brook 1"]; James Brook, *The Use of Statistical Evidence of Identification in Civil Litigation: Well-Worn Hypotheticals, Real Cases, and Controversy*, 29 ST. LOUIS U.L.J. 293 (1985) [hereinafter "Brook II"]; L. Jonathan Cohen, *Subjective Probability and the Paradox of the Gatecrasher*, 1981 ARIZ. ST. L.J. 627 [hereinafter "L.J. Cohen II"]; Michael Finkelstein & William Fairley, *A Bayesian Approach to Identification Evidence*, 83 HARV. L. REV. 489 (1970); John Kaplan, *Decision Theory and the Fact-Finding Process*, 20 STAN. L. REV. 1065 (1968); David Kaye, *The Laws of Probability and the Law of the Land*, 47 U. CHI. L. REV. 34 (1979); Charles Nesson, *Reasonable Doubt and Permissive Inference: The Value of Complexity*, 92 HARV. L. REV. 1187 (1979) [hereinafter "Nesson I"]; Charles Nesson, *The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts,* 98 HARV. L. REV. 1357 (1985) [hereinafter "Nesson II"]; Michael Saks & Robert Kidd, *Human Information Processing and Adjudication: Trial by Heuristics*, 15 LAW & SOC'Y REV. 123 (1980); and Lawrence Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329 (1971).

2 Base rates describe the frequency with which a relevant attribute occurs among members of a reference population. A base rate may also be thought of as the probability that a randomly selected member of a reference population will have the relevant attribute.

3 *See, e.g.,* Finkelstein & Fairley, *supra* note 1; Kaye, *supra* note 1.

irrelevant absent a causal link to the specific case before the jury;[4] yet others contend that it may be relevant but is insufficient to support a verdict.[5]

The second issue concerns the use of subjective probabilities, or personal estimates of the probability that a factual proposition is true. Commentators dispute the feasibility and appropriateness of encouraging such estimates and employing probability theory to assess the combined significance of several subjective estimates that pertain to the same case.[6]

The third issue, which embraces the first two, is the underlying purposes trials should serve. It is widely agreed that an important objective of trials is verdict accuracy, or attempting to impose guilt or liability when, but only when, the actual facts warrant it.[7] Yet trials may serve other objectives as well, thus inevitably complicating the choice of appropriate rules and procedures.

One cannot determine the merits of using base rate evidence and subjective probabilities without stipulating the underlying purposes in structuring trials. Yet commentators have not always distinguished clearly between verdict accuracy and what we will term "policy concerns" (*i.e.*, policies distinct from verdict accuracy that are implicated by trials).[8] Instead, they often have let their views about one dictate their views about the other. Since the two are conceptually distinct, however, it is useful to consider separately the implications of probabilistic evidence for each in order to identify and evaluate the tradeoffs that may be involved.

This Article will perform the first half of the necessary analysis by examining the effect of overtly probabilistic evidence and methods on verdict accuracy. We will show that overtly probabilistic evidence is no less probative of legally material facts than other types of evidence. We will suggest, moreover, that rules of probability theory such as Bayes' theorem can improve the accuracy with which juries evaluate evidence in particular cases, even when applied (where feasible) to subjective probabilities.

Our analysis will not reach the question of whether, on balance, greater use of overtly probabilistic evidence and methods at trial is desirable. This determination depends on the value attached to specific policy concerns other than verdict accuracy. We hope to show, however, that a refusal to employ overtly probabilistic evi-

---

4   *See, e.g.,* L.J. COHEN I, *supra* note 1; L.J. Cohen II, *supra* note 1.

5   *See, e.g.,* Nesson I, *supra* note 1; Nesson II, *supra* note I; Tribe, *supra* note 1.

6   *See, e.g.,* L.J. COHEN I, *supra* note 1; Brilmayer & Kornhauser, *supra* note 1; L.J. Cohen II, *supra* note 1; Tribe, *supra* note I.

7   *See, e.g.,* Saks & Kidd, *supra* note I; Tribe, *supra* note 1.

8   While using the term "policy concerns" in this fashion for convenience, we recognize that verdict accuracy is a policy like any other.

dence and methods has a cost, namely, increasing the probability of inaccurate verdicts. At times, this cost may be worth incurring if the expected benefits are sufficient; but it should not be denied or ignored.[9]

Section I of this Article discusses the basic terms involved in a consideration of overtly probabilistic evidence. Section II considers the significance for verdict accuracy of using base rate evidence. Section III discusses the use of subjective probabilities. Section IV discusses problems with the use at trial of probabilistic evidence and methods. Section V provides a summary and conclusion.

I

DISCUSSION OF TERMS

This section discusses the basic terms that govern our consideration of overtly probabilistic evidence. It begins by defining verdict accuracy and distinguishing that goal from other policy concerns of the trial process. It then defines overtly probabilistic evidence, including base rates and subjective probabilities. Finally, it describes certain basic rules of probability theory, such as the product rule and Bayes' theorem.

A. Objectives of the Trial Process

1. *Verdict Accuracy*

At many trials, both civil and criminal, one of the parties is factually in the right. A defendant either was or was not correctly identified as the murderer; a plaintiff either did or did not make statements expressing oral agreement to a contract. In such cases, if all the relevant facts became known after trial, it would be possible to decide whether or not the verdict was factually accurate.

Of course, some cases are more ambiguous. Was the defendant in a murder trial legally insane? Did a tobacco company that is sued by a smoker's estate act reasonably in light of the information it possessed about the medical effects of smoking? Even in these cases, however, subsidiary questions of disputed fact may have correct answers. In such cases, if one knows the details of the fact-finder's reasoning, one can speak of the factual accuracy of its judgments with respect to these questions.

---

9   We also will not attempt to quantify this cost or to determine when the expected benefits of overtly probabilistic evidence and methods are sufficiently high to mandate their use. Yet it is worth noting that scientific advances may cause the usefulness of such evidence and methods to increase over time. For example, improved methods for testing blood, hair, and skin samples found at the scene of a crime may increase the availability of statistical evidence regarding the percentage of the population that, like the defendant, would be possible sources of such samples.

To the extent that the facts of a case, if known to the fact-finder, would require a particular verdict in light of the applicable law, one can speak of *verdict accuracy*. In practice, the criteria for measuring verdict accuracy generally do not exist; one often cannot be certain about all of the relevant disputed facts. Yet the lack of a truth criterion does not mean that verdicts are neither accurate nor inaccurate; it merely indicates limits to our knowledge about particular cases.

Verdict accuracy is one of the principal goals of the trial process.[10] Even in the absence of separate policy concerns that influence the conduct of trials, however, accuracy cannot be guaranteed. Gaps and mistakes in fact-finding inevitably will occur in some cases and thus lead to inaccurate verdicts. Given this problem, along with the lack of a truth criterion even after trial, the best that can be accomplished in relation to verdict accuracy is to minimize the number of inaccurate verdicts that one reasonably expects.

A fact-finder who was motivated solely by verdict accuracy would always hold in favor of whichever party appeared more likely to be correct in light of the evidence. Assuming a positive correlation between the fact-finder's belief and the true state of affairs, such a verdict would minimize the likelihood of individual case error, and thus promote verdict accuracy in general. This objective might suggest following an overtly probabilistic methodology: for example, attempting to determine mathematically, based on one's views about particular issues, whether one believes that the likelihood of the plaintiff's being entitled to recover is greater than fifty percent.

2. *Policy Concerns Apart From Verdict Accuracy*

Verdict accuracy is not and should not be the only objective served by trials. If it were, one might decry not only all testimonial privileges but even the legal system's refusal to countenance the use of interrogation tools such as torture and sodium pentothal.[11] However, since verdict accuracy is the goal that we isolate for analysis in this Article, it is useful to distinguish certain other policy concerns that we are expressly disregarding.

Many policy concerns relate to the trial (or pretrial) process, rather than to the verdict reached in a particular case. For example, consider the fourth amendment ban on certain searches and seizures, or the wide array of testimonial privileges that can be asserted to resist disclosure of confidential communications. Such rules exclude probative evidence, and thus presumably reduce verdict accuracy. They do so to protect certain individual rights or pro-

---

[10]   "The basic purpose of a trial is the determination of truth ...." Tehan v. United States *ex rel.* Shott, 382 U.S. 406, 416 (1966).

[11]   *See* Kaye, *supra* note 1, at 45 n.41.

mote particular business or personal relationships. In practice, some of these rules may promote particular outcomes (*e.g.*, acquittal of criminal defendants in the case of the fourth amendment). Yet this is in a sense incidental; the rules are based on concern about means rather than ends.

Other policy concerns relate directly to the verdict reached by the fact-finder. For example, consider the requirement that criminal guilt be proven beyond a reasonable doubt. This requirement reflects the view that an erroneous conviction is far more regrettable than an erroneous acquittal.[12]

Another possible policy concern that relates directly to verdicts involves the fact-finder's confidence about the probability that particular facts are true. Consider, for example, two civil trials in which the jury believes that there is a sixty percent chance that the plaintiff would be entitled to recover if all of the facts were accurately known. In the first case, the jury has heard extensive evidence and believes it has a very strong grasp of the issues. It therefore believes that additional evidence would be unlikely to change its assessment significantly.

In the second case, however, the jury has heard only scant evidence and has little confidence in its assessment of the probabilities. It already has drawn all reasonable inferences from the evidence (and lack thereof) in determining that the plaintiff is sixty percent likely to be entitled to recover. Yet it is concerned that additional evidence, if made available, might significantly shift its view of the probabilities.

The jury in the second case is experiencing uncertainty about its uncertainty, or second-order uncertainty.[13] It is subject, not only to the first-order uncertainty implicit in the sixty percent probability, but also to uncertainty about this very estimate. The jury views its estimate as unresilient. As far as it knows, only a small amount of additional information might substantially change its estimate of the probability that the plaintiff is correct.

Empirical studies suggest that people making judgments and decisions generally take both first-order and second-order uncertainty into account.[14] For example, people engaged in gambling

---

12   The reasonable doubt requirement also reflects policy concerns, such as making less visible, obvious, or seemingly intentional the occasional conviction of an innocent defendant. *See* Daniel N. Shaviro, *Statistical Probability Evidence and the Appearance of Justice*, 103 HARV. L. REV. 530 (1989).

13   *See* Hillel Einhorn & Robin Hogarth, *Ambiguity and Uncertainty in Probabilistic Inference*, 92 PSYCHOLOGICAL REV. 433 (1985).

14   *See id.; see also* Selwyn Becker & Fred Brownson, *What Price Ambiguity? Or the Role of Ambiguity in Decision Making*, 72 J. POL. ECON. 62 (1964); Frank Yates & Lisa Zukowski, *Characterization of Ambiguity in Decision Making*, 21 BEHAV. SCI. 19 (1976).

will often sacrifice expected payoff value in return for decreases in second-order uncertainty.[15] Some commentators have suggested that second-order uncertainty about one's estimate of the probability of guilt or liability suggests declining to allow a verdict for the prosecutor or plaintiff, even if this estimate exceeds an otherwise appropriate threshold (*e.g.*, 50.1 percent in a civil case).[16]

If these commentators are correct, it should be clear that they are voicing policy concerns apart from verdict accuracy. Just as it would be irrational to be influenced by second-order uncertainty if one's sole objective as a bettor is to maximize expected profit,[17] so it is a mistake to believe that verdict accuracy can be enhanced by second-order considerations.

B.   Overtly Probabilistic Evidence

All evidence is probabilistic, in the sense that there is a risk of error in relying on it to support a factual conclusion about a case.[18] Overtly probabilistic evidence, however, makes the risk of error explicit.

1.   *Base Rates*

Perhaps the most common type of overtly probabilistic evidence involves base rates. A base rate may be defined as the relative frequency with which an event occurs or an attribute is present in a population. The base rate for an event or attribute equals the probability that it will be present in any randomly selected member of the reference class prior to the introduction of case-specific or individuating information.

As an example of a base rate, consider a hypothetical developed by Professor Nesson in which there are twenty-five prisoners in an enclosed yard, twenty-four of whom collaborate in the murder of a prison guard. The twenty-fifth prisoner hides in a nearby shed during the crime and is innocent of all involvement. Prior to the introduction of any evidence about which of the twenty-five prisoners did or did not participate in the murder, the probability that a randomly

---

15   Daniel Ellsberg, *Risk, Ambiguity, and the Savage Axioms*, 75 Q.J. ECON. 643 (1961).

16   *See, e.g.*, L.J. COHEN 1, *supra* note 1; Brilmayer, *supra* note 1; L.J. Cohen II, *supra* note 1.

17   Bettors may be rationally deterred by second-order uncertainty if they have betting objectives other than profit maximization. For example, one may dislike feeling foolish if one not only loses a bet but discovers that additional information would have changed one's estimates sharply. Or one may prefer betting on certain sixty percent probabilities, rather than on uncertain sixty percent probabilities, due to an aversion to increasing variance in long run success rates. *See* Bruno De Finetti, *Probabilities of Probabilities: A Real Problem or a Misunderstanding?*, in NEW DEVELOPMENTS IN THE APPLICATIONS OF BAYESIAN METHODS (Ahmet Aykac & Carlo Brumat eds. 1977).

18   Tribe, *supra* note 1, at 1330 n.2.

selected prisoner is guilty is 24/25, or ninety-six percent.[19]

In some instances (such as the prisoner case), a base rate frequency may express the probability of ultimate guilt or liability. In practice, however, this is extremely rare. The fact to which a base rate relates may be only one of several at issue in a case. One example is a case in which the base rate helps determine who committed an act, but liability also requires a finding of negligence. Moreover, there may be evidence apart from the base rate that relates to the same fact (*e.g.*, conflicting eyewitness testimony about identity). Thus, base rate evidence often must be combined with other evidence and inferences before a probabilistic conclusion can be reached.

2. *Subjective Probabilities*

Another type of overtly probabilistic estimate arises when one makes a personalistic or subjective assessment of the probability that a particular claim is true. For example, in the prisoner hypothetical, following the introduction of additional evidence a juror might believe that there is a seventy-five percent chance that the defendant participated in the murder.

One could try to describe the juror's belief as suggesting that, if a case with identical evidence took place one hundred times, then the juror would expect that in seventy-five cases the defendant actually participated in the murder. The problem with so expressing the personal probability estimate, however, is that the set of evidence in this case may be unique. The prisoner either is or is not actually guilty, and the set of evidence cannot be tested one hundred times.

Accordingly, in conventional discourse, the juror's estimate describes only a state of mind or degree of belief. Operationally speaking, a seventy-five percent subjective estimate of guilt means that the juror would be as willing to wager on guilt as to wager (for example) that a card randomly selected from a well-shuffled deck would be a spade, club, or diamond, rather than a heart.[20]

C. Probabilistic Methodologies

The rules of mathematics often provide means of assessing the combined significance of more than one probability estimate. For example, if the likelihood of Facts X, Y, and Z all are known, and enough is known about how the truth of each fact affects the truth of

---

19   *See* Nesson I, *supra* note 1, at 1192-93.
20   *See* LEONARD J. SAVAGE, FOUNDATIONS OF STATISTICS (1954); Tribe, *supra* note 1, at 1346-49. The juror is assumed for this purpose to be a profit-maximizer who is undeterred by second-order uncertainty.

the others, then one can calculate the likelihood that any combination of the three is true.

It is widely recognized that base rates and subjective probabilities obey the same mathematical rules.[21] Thus, the rules can be used to assess the combined significance of subjective probabilities, yielding revised probabilities that are internally consistent.

I. *The Product Rule*

One application of mathematics to multiple probabilities is the product rule, which states that the probability of the joint occurrence of Events A and B is (1) the probability of Event A, times (2) the probability of Event B given the occurrence of Event A.[22] When the events in question are independent,[23] the probability of their joint occurrence is simply the product of their individual probabilities.[24] For example, if the chance of any coin toss yielding heads is $1/2$, then the chance that any two coin tosses will both be heads is $1/2 \times 1/2$, or $1/4$.

Two events may also be completely dependent. Thus, assume that Events A and B each have a $1/2$ chance of occurring, but that Event B occurs if and only if Event A occurs. In such a case, the chance that both A and B will occur is equal to the chance of either occurring, or $1/2$.[25]

In some cases, events may be less completely dependent. The occurrence of Event A may make Event B more (or less) likely, yet it may be possible for either event to occur when the other does not. In such cases of partial dependence, the probability that two events both will occur is somewhere between the lower bound established by the independence form of the product rule and the higher bound that follows from complete dependence. Thus, in the last example above, partial dependence would suggest that the chance of Events A and B both occurring is somewhere between $1/4$ and $1/2$.[26]

---

[21] The point was originally made by Savage. *See* Kaye, *supra* note 1, at 44 n.35.

[22] *See, e.g.,* RICHARD A. WEHMHOEFER, STATISTICS IN LITIGATION 39 (1985); Allison Cullison, *Probability Analysis of Judicial Fact-Finding: A Preliminary Outline of the Subjective Approach*, 1 U. TOL. L. REV. 538, 541-42 (1969).

[23] *See* People v. Collins, 68 Cal. 2d 319, 438 P.2d 33, 66 Cal. Rptr. 497 (1968) (prejudicial error committed by trial court in admitting expert testimony on the joint probability of a series of events without first establishing the independence of those events).

[24] Two Events, A and B, are independent if the probability of A equals the probability of A given that B has occurred, and the probability of B equals the probability of B given that A has occurred (*i.e.*, $P(A) = P(A/B)$ and $P(B) = P(B/A)$).

[25] Formally, if $P(A) = 1/2$, $P(B) = 1/2$, $P(A/B) = 1$, and $P(B/A) = 1$, then $P(A \text{ and } B) = P(A/B) \times P(B) = P(B/A) \times P(A) = 1 \times 1/2 = 1/2$.

[26] In light of the attack on probabilistic methods in *Collins*, 68 Cal. 2d 319, 438 P.2d 33, 66 Cal. Rptr. 497, the difference between the independence form of the product rule and the general rule is worth emphasizing. In *Collins*, a series of estimated probabilities

2. *Bayes' Theorem*

Bayes' theorem is a method for combining unconditional and conditional probabilities in the face of new evidence to derive revised estimates of the probability that a particular claim is true.[27] To illustrate Bayes' theorem, consider the following problem posed by Tversky and Kahneman:

> A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data: (i) 85% of the cabs in the city are Green and 15% are Blue. (ii) A witness identified the cab as a Blue Cab. The court tested his ability to identify cabs under the appropriate visibility conditions. When presented with a sample of cabs (half of which were Blue and half of which were Green) the witness made correct identifications in 80% of the cases and erred in 20% of the cases. Question: What is the probability that the cab in-

---

were multiplied to determine the probability of their joint occurrence, based on the unsupported assumption that all of the probabilities were independent. Thus, as indicated by the California Supreme Court and numerous subsequent commentators, the product rule was misused in *Collins. See, e.g.,* MICHAEL O. FINKELSTEIN, QUANTITATIVE METHODS IN LAW (1978); William Fairley & Frederick Moesteller, *A Conversation About Collins,* 41 U. CHI. L. REV. 242 (1974); Tribe, *supra* note 1. The validity of the product rule itself generally is not disputed, although L.J. COHEN I, *supra* note 1, at 32, provides an alternative probability calculus.

[27] Bayes' theorem follows directly from the product rule discussed earlier, which holds that the joint probability of two events, H and E, equals the product of the conditional probability of one of the events given the second event, and the probability of the second event. In mathematical notation:

$$P(H \ \& \ E) = P(H/E) \ P(E)$$
$$P(H \ \& \ E) = P(E/H) \ P(H)$$

Equating the two right sides of the equations gives:

$$P(H/E) \ = \ \frac{P(E/H) \ P(H)}{P(E)}$$

where $P(E) = P(E/H) \ P(H) + P(E/-H) \ P(-H)$ for binary hypotheses.

Bayes' theorem may also be expressed in odds form as follows:

$$\frac{P(H/E)}{P(-H/E)} \ = \ \frac{\dfrac{P(E/H) \ P(H)}{P(E)}}{\dfrac{P(E/-H) \ P(-H)}{P(E)}} \ = \ \frac{P(E/H) \ P(H)}{P(E/-H) \ P(-H)} \ = \ \frac{P(H)}{P(-H)} \ \times \ \frac{P(E/H)}{P(E/-H)}$$

The letters H and E may be thought of as standing for Hypothesis and Evidence respectively. P(H) and P(−H) refer to the likelihoods of the truth and falsity of hypothesis H prior to the collection of additional evidence. In Bayesian terminology, P(H) and P(−H) are "priors" and their ratio is the "prior odds." P(E/H) and P(E/−H) represent the information value of the evidence if H is true and if H is false respectively; their ratio is the "likelihood ratio." P(H/E) and P(−H/E) are the likelihoods that the hypothesis is true and false in light of the evidence; their ratio is the "posterior odds," which represents the combination of the prior odds and the likelihood ratio.

CORNELL LAW REVIEW [Vol. 75:247

volved in the accident was Blue rather than Green?[28]

The answer to the question (assuming no other evidence) under Bayes' theorem may be grasped intuitively by considering what would happen in one hundred cases where the above problem arose, assuming a distribution of events in accord with the probabilities provided. First, based on the percentages given in (i), in eighty-five cases the cab would be Green. Given the witness's eighty percent accuracy rate, in sixty-eight of these cases he would correctly state that the cab was Green, and in seventeen cases he would incorrectly state that the cab was Blue. Second, in fifteen cases the cab would be Blue. Given the witness's accuracy, in twelve of these cases he would correctly state that the cab was Blue, and in three cases he would incorrectly state that it was Green.

In the present case, however, the witness stated that the cab was Blue. Accordingly, the two relevant frequencies above are the twelve cases in which he said it was Blue when it indeed was Blue, and the seventeen cases in which he said it was Blue when it actually was Green. Thus, the probability that the cab in fact was Blue, given that the witness said it was Blue, is only 12/29, or about forty-one percent.

As Kahneman and Tversky demonstrate, this answer is somewhat counter-intuitive. People tend to focus disproportionately on the witness's eighty percent accuracy rate, and to undervalue (or altogether ignore) the base rate information concerning the frequency of the two types of cabs.[29] They do not seem to appreciate that, because the witness is much more likely to have encountered a Green cab than a Blue cab, he would have more opportunities to err about the former than to perceive the latter accurately.

By combining prior probabilities and likelihood values to determine posterior probabilities, Bayes' theorem provides a strategy for updating one's probability estimates as more information becomes available. Of course, the theorem has nothing to say about the veracity or reliability of the input probabilities. A posterior probability value is no more valid than the inputs that contribute to its computation. Yet Bayes' theorem does ensure a posterior value that is consistent with a decisionmaker's other probabilistic estimates.

---

28 Amos Tversky & Daniel Kahneman, *Causal Schellas in Judgments Under Uncertainty*, in PROGRESS IN SOCIAL PSYCHOLOGY (Martin Fishbein ed. 1980).

29 Since Tversky and Kahneman introduced the cab problem in 1972, numerous studies have indicated that people apparently pay relatively little attention to the experimenter-supplied base rate information. *See* Maya Bar-Hillel, *The Base Rate Fallacy Controversy*, in DECISION MAKING UNDER UNCERTAINTY (R. Scholtz ed. 1983).

II
VERDICT ACCURACY AND THE USE OF BASE RATE EVIDENCE

Debate about base rate evidence has largely been dominated by hypotheticals, or perhaps by a single hypothetical that appears in several different guises. One form of the hypothetical is Professor Nesson's prisoner case, in which twenty-four out of twenty-five prisoners participated in the murder of a prison guard. Nesson describes the ninety-six percent probability that a randomly selected prisoner participated in the murder as "purely statistical" evidence.[30] He argues that the case would not and should not be allowed to go to the jury absent case-specific evidence pertaining to the particular defendant on trial.[31]

Two related hypotheticals that concern civil rather than criminal liability have also been widely discussed in the literature. First, in the blue bus case described by Professor Tribe, a plaintiff is hit by a bus that she believes was blue, but otherwise cannot describe. She offers no evidence that the defendant bus company was liable, beyond establishing that it operates eighty percent of the blue buses in town. She is not allowed to recover.[32]

Second, in British philosopher L.J. Cohen's "gatecrasher paradox," 1000 people whose identities are known attend a rodeo to which the promoter sold only 499 tickets. Thus, it appears that 501 attendees, or slightly more than half, entered without paying, but they cannot be identified except as members of the larger group. Here too the base rate evidence, suggesting a 50.1 percent probability that a particular rodeo attendee was a gatecrasher, is described as purely statistical, and thus as insufficient to support a verdict.[33]

The prisoner, blue bus, and gatecrasher hypotheticals share two critical characteristics. First, they provide a base rate for guilt or liability, not simply for a factual characteristic distinct from the ultimate legal issue. Second, they provide no evidence other than a base rate (except insofar as inferences can be drawn from the parties' litigating position or their failure to introduce more evidence).[34]

---

30   Nesson I, *supra* note 1, at 1193.

31   *Id.*

32   Tribe, *supra* note 1, at 1340-41 (loosely based on the real case of Smith v. Rapid Transit, Inc. 317 Mass. 469, 58 N.E.2d 754 (1945)).

33   L.J. COHEN I, *supra* note 1, at 75.

34   *See* Neil B. Cohen, *Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge*, 60 N.Y.U. L. REV. 385 (1985) [hereinafter "N. Cohen I"]; Kaye, *supra* note 1.

These characteristics are highly unusual; indeed, so much so that it is far from clear that existing law requires the suggested verdicts.[35] By being so unusual, the hypotheticals divert attention from issues more commonly posed by the use of base rate evidence—issues such as the need to combine such evidence with additional, more subjective evidence, or to choose among different base rates.

Perhaps the hypotheticals' main advantage is that they dramatize the anti-Bayesian position: that base rate evidence is insufficient to support a verdict, and that this insufficiency results from its supposed inferiority to evidence that is considered specific to the case at hand.[36] This Section examines these premises from the standpoint of verdict accuracy, considering why and when base rate evidence is relevant, as well as whether it is less probative than case-specific evidence. It then considers the use of probabilistic methodologies with respect to base rate evidence.

A. The Relevance of Base Rate Evidence

1. *Why Base Rates Are Relevant*

Base rate evidence is relevant because it can assist one in assessing whether a particular fact is true. Consider again the prisoner hypothetical. The base rate of ninety-six percent does not, of course, establish with certainty whether any particular defendant is guilty. However, it provides the long run accuracy rate that would result from repeatedly holding for the prosecution (assuming random selection of prisoner defendants from among the group). The conviction of all twenty-five prisoners would lead to twenty-four accurate verdicts and one inaccurate verdict.

It may be objected that long run verdict accuracy is not the proper goal in any particular case. Instead one should seek to decide *that* case accurately.[37] Some have based this view on policy concerns apart from verdict accuracy, such as an asserted ethical need to treat people as individuals rather than as sample members of larger reference classes.[38]

From the standpoint of verdict accuracy, however, the distinc-

---

35 *See* Brook II, *supra* note 1, at 294, 299-310; D.H. Kaye, *The Admissibility of "Probability Evidence" in Criminal Trials Part II*, 27 JURIMETRICS J. 160 (1987); D.H. Kaye, *The Admissibility of "Probability Evidence" in Criminal Trials Part I*, 26 JURIMETRICS J. 343, 343-46 (1986).

36 *See, e.g.*, Brilmayer, *supra* note I, at 675-76; Nesson II, *supra* note I, at 1357, 1379.

37 *See, e.g.*, Brilmayer & Kornhauser, *supra* note 1, at 150.

38 *See, e.g., id.* at 150 n.119; Barbara D. Underwood, *Law and the Crystal Ball: Predicting Behavior with Statistical Inference and Individualized Judgment*, 88 YALE L.J. 1408 (1979). Yet, one might wonder whether a person who lost an important legal case because of the exclusion of base rate evidence would derive much solace from being treated as a unique human being.

tion between the long run and the particular case is invalid. Since the long run is composed of a series of individual cases, information (such as base rates) that informs long-run accuracy rates must likewise inform the likelihood of accuracy in individual cases.

In an example like the prisoner hypothetical, where there is no evidence differentiating the members of the reference class from which the base rate is derived, all members share the same probability of possessing the measured attribute (*i.e.*, ninety-six percent). Only when differentiating evidence becomes available does it make sense to regard an individual member's probability as anything other than the base rate frequency for the general reference class to which he belongs.

Even given information differentiating the members of a reference class, base rate evidence remains relevant to specific cases. For example, in the prisoner hypothetical, assume that a particular defendant has an unusually peaceful disposition. This fact may suggest that the ninety-six percent base rate overstates the probability of his guilt. Yet the character evidence does not make this base rate evidence irrelevant. One's posterior probability estimate should and likely would be very different than if, say, only one of the twenty-five prisoners was guilty.

2. *Assessing the Relevance of Particular Base Rates*

a. *Base Rate Specificity*

Even if the relevance of base rates in general is assumed, there is some question as to *which* base rates should be used in particular cases. This problem tends to be disguised by hypotheticals such as the gatecrasher and prisoner cases. In such cases, even if one is reluctant to ground a verdict on the base rates alone, it seems clear that the base rates were derived from appropriate reference classes (*i.e.*, the number of rodeo attendees or prisoners in the yard).

In Kahneman and Tversky's cab problem,[39] however, one may question the appropriateness of the reference class (*i.e.*, cabs in the city where the accident occurred). This reference class is relatively unspecific; it reflects fewer seemingly important features of the case at hand than would the reference class "cabs in the city that were involved in accidents," or "cabs in the city that were involved in accidents at night."

As noted earlier, research shows that people ignore or attach little weight to the citywide base rate in assessing the probabilities in the cab problem, apparently considering it insufficiently informative by reason of its unspecificity. When the reference class is narrowed

---

[39] *See supra* text accompanying note 28.

from cabs in the city to cab accidents in the city, they accord it more weight, although still less than that suggested by Bayes' theorem.[40]

L.J. Cohen argues that people are correct to ignore frequencies as unspecific as the citywide base rate for cabs:

> [W]hy on earth should it be supposed that subjects, asked to esti-
> mate the unconditional probability that the cab *involved in the acci-*
> *dent* was [B]lue, ought to take into account a prior distribution of
> colours that would at best be relevant only if the issue at stake was
> just about the colour of a cab that was said to have been seen
> somewhere, *not necessarily in an accident*, and was taken to be
> [B]lue?[41]

Cohen concludes that the citywide cab distribution of eighty-five percent Green and fifteen percent Blue provides a "very weak foundation for an estimate of the *relevant* base rate."[42]

Yet what is meant by the *relevant* base rate (or the *relevant* reference class)?[43] The phrases imply exclusivity. However, while "cabs in accidents in the city" is a more informative reference class than "cabs in the city"—since it takes into account an additional factor of potentially great significance (namely, the propensity of different cab companies to get into accidents)—it is not even the most informative reference class. For example, "cabs in accidents in the city *at night*" would take into account an additional factor of potential importance. If location within the city is important, a still better reference class might be "cabs in accidents at night in the same general area"—or perhaps even "cabs in accidents at night at exactly the same location."[44]

It would seem, then, that there is no such thing as *the* relevant base rate. Some base rates are better than others in that they are derived from reference classes that take into account more potentially important information from the particular case.[45] An increase

---

40   Tversky & Kahneman, *supra* note 28, at 157-58.
41   L. Jonathan Cohen, *Are There Any A Priori Constraints on the Study of Rationality?*, 4 BEHAV. & BRAIN SCI. 359, 365 (1981) (emphasis in original).
42   *Id.* (emphasis in original).
43   *See* L.J. Cohen II, *supra* note 1, at 633. (L.J. Cohen discusses the "suitably narrowed down reference class" that should be used to decide a case).
44   As the reference class is increasingly refined, however, the sample space from which the base rate is computed becomes smaller. This may be problematic because base rates derived from small sample spaces are less reliable than those derived from larger sample spaces. *See infra* note 48 and accompanying text; *see also* Kevin Lanning, *Some Reasons for Distinguishing Between "Non-normative Response" and "Irrational Decision"*, 121 J. PSYCHOLOGY 109, 112 (1987) (discussion of the "trade-off between fineness of analysis and statistical power" when choosing reference classes).
45   The priority between base rates that reflect different information is not always clear. Thus, in the cab problem, people might disagree about whether a base rate for cabs in accidents in the city, or for cabs operating at night, is more informative. Where this is the case, presumably both base rates should be admitted, and the jury allowed to

in base rate specificity generally reduces second-order uncertainty, since it reduces the amount of unknown information that might change one's estimate. Yet a base rate need not meet any particular standard of specificity in order to be relevant.

From the perspective of verdict accuracy, it is unjustifiable to ignore, by reason of its unspecificity, the best available base rate. For example, to ignore the citywide base rate for Blue and Green cabs, when no better base rate is available, is tantamount to assuming that the likelihood of Blue and Green cabs being involved in an accident is the same (*i.e.*, fifty percent for each).[46] However, the citywide base rate provides some evidence that this is not the case. Since eighty-five percent of the cabs in the city are Green, one's best gness—in the absence of any information to suggest otherwise[47]—is that eighty-five percent of the cabs in accidents are Green. Likewise, one's best guess is that eighty-five percent of the cab accidents at night involve Green cabs, and so on for further refinements of the cab base rate.

In conclusion, (1) in many cases, no particular base rate or reference class is uniquely relevant, (2) base rates derived from relatively unspecific reference classes may be associated with greater second-order uncertainty, and (3) in the absence of grounds for modifying a computed frequency, a base rate's specificity should have no effect on its evidentiary weight from the standpoint of verdict accuracy.

b. *Base Rate Sampling Problems*

A base rate can reflect an examination of either all members of a reference class or a sample of members. When based on a sample, its use requires the inference that the sample base rate accurately represents that for the entire reference class.

The use of non-random or small samples generally promotes second-order uncertainty by creating or increasing doubt that the

---

assess their comparative relevance. However, when it is clear that one base rate is more specific than another, in that it incorporates the same information plus additional information, then, provided statistical reliability concerns are satisfied, perhaps only the more specific one should be admitted. The less specific base rate arguably would be excludable under FED. R. EVID. 403, on the ground that its limited probative value is outweighed by the danger of confusion.

46   The odds form of Bayes' theorem indicates that prior probabilities have no influence on posterior probabilities only when the prior odds ratio, $P(H)/P(-H) = 1$. *See supra* note 27. In cases such as the cab problem, where there are two mutually exclusive and exhaustive hypotheses, this ratio is 1 only when the prior probability of each hypothesis is 50 percent.

47   An example of information suggesting otherwise would be an inference of strategic behavior by the party that introduced the citywide base rate, if it seems plausible that the party could have computed a more specific base rate without undue difficulty.

sample accurately represents the reference class as a whole. By contrast, when sufficiently large random samples are used, second-order uncertainty is reduced and one may be confident that sample and true reference class population base rates are similar.

However, the mere presence of second-order uncertainty does not inform one about the direction in which a sample base rate should be adjusted. Such adjustments should be made only when other information (including prior probability estimates) is available.

Accordingly, while large samples (like more specific reference classes) are generally preferable because they incorporate more information, verdict accuracy concerns do not support treating them differently from small samples in the absence of grounds for modifying computed frequencies.[48] However, given additional information, the conclusions derived from small samples should be more readily revised than those derived from large samples.

In practice, however, the use of evidence based on small samples may often pose serious problems, especially in the absence of reasonable cause for not using evidence based on larger samples. When a litigant introduces a base rate derived from a small sample without good cause, the inference of strategic behavior may be so strong as to overwhelm the evidentiary value of the base rate.

3. *Comparison of Base Rate Evidence to Case-Specific Evidence*

Despite the widely conceded relevance of base rate evidence, such evidence is commonly regarded as inferior to other kinds of evidence. Base rate evidence is not always disparaged by reason of its use of numbers, which after all "convey information in a different form, but not of a different kind."[49] Instead, it is often disparaged by reason of its derivation from background data, comprised of the aggregate characteristics of a reference class. Background evidence is considered somehow inferior to evidence that is individuating and specific to the case at hand.

The intuitive basis for the distinction between background and case-specific evidence is described by Professors Brilmayer and Nesson (neither of whom argues that the distinction is entirely correct on its face). With regard to the blue bus and gatecrasher hypotheti-

---

48   Professor Neil Cohen apparently would treat small samples differently from large ones, since he argues that confidence intervals, which capture second-order uncertainty, should be incorporated into the burden of proof. *See* N. Cohen I, *supra* note 34, at 404-09; and Neil B. Cohen, *Conceptualizing Proof and Calculating Probabilities: A Response to Professor Kaye*, 73 CORNELL L. REV. 78, 85 (1987). As Neil Cohen has come to recognize, his views reflect a combination of verdict accuracy and other policy concerns which do not necessarily minimize what he terms "fact errors."

49   Brook II, *supra* note 1, at 294.

cals, Brilmayer notes that base rate evidence cannot tell us that the ABC company "really" was at fault because it operated so many blue buses, or that Sally Smith "really" was a gatecrasher. Instead, she states, "[t]here is only a background statistic about the number of buses owned, or the number of tickets sold."[50]

As Brilmayer acknowledges, however, the same can be said about case-specific evidence.[51] After all, both epistemologically and in most cases practically, how can one ever "really" know anything? For example, how can eyewitness testimony convince us that Sally Smith "really" was a gatecrasher, rather than that she was probably one? All evidence is probabilistic, requires inferences to support an ultimate conclusion, and involves a risk of error if thought to establish that conclusion.[52] Base rate evidence is different only in that it makes these uncertainties explicit.

Nesson further describes the underlying belief in the inferiority of background statistical evidence through a hypothetical in which one considers whether a card drawn from a well-shuffled deck was a king. Absent case-specific evidence, statistical background evidence suggests that the chances are 48/52 that it was not a king. Nesson suggests that, if we rely on the background evidence but the card turns out to be a king, we may conclude, not that we wrongly assessed the probabilities, but that an improbable event occurred.[53]

Nesson contrasts this hypothetical to a second one in which, while perhaps not knowing anything about the composition of the deck of cards, we have case-specific evidence in the form of a quick glance at the card which suggests that it is not a king. Here, if we rely upon our eyesight but the card turns out to be a king, we may conclude that we were wrong.[54]

As Nesson appears to recognize, the distinction is only semantic.[55] In the second hypothetical, we presumably realize (or should realize) that we usually, but not always, see an object correctly when we are permitted a quick glance. We rely on the quick glance, absent better information, because it is correct most of the time. Thus, the second hypothetical, like the first one, merely involves an improbable event, rather than a decisional error.

---

50   Brilmayer, *supra* note 1, at 675.
51   *Id.* at 676.
52   Brilmayer nonetheless claims that the legal system is correct in treating background evidence as inferior to case-specific evidence, but she describes her position as a "second-order principle," presumably reflecting policy considerations. *See id.*
53   Nesson II, *supra* note 1, at 1361-62.
54   *Id.*
55   Nesson I, *supra* note 1. On the purely semantic nature of the distinction between background and case-specific evidence, see Saks & Kidd, *supra* note 1, at 151-54; Richard Schmalbeck, *The Trouble With Statistical Evidence*, 49 LAW & CONTEMP. PROBS. 221, 232-33 (1986).

Nesson views the semantic distinction between background sta-
tistical evidence and case-specific evidence as relevant because he
believes that people consider it so. He argues that, if people feel the
need for an evidentiary "bridge" from the general to the specific,
they will respond differently to verdicts based on the two types of
evidence.[56] Thus, he objects to base rate evidence based on policy
concerns, while recognizing that his views involve a sacrifice of ver-
dict accuracy.[57]

From the standpoint of verdict accuracy, the equivalence be-
tween background and case-specific evidence is difficult to dis-
pute.[58] Even cases involving "naked statistical evidence" (*i.e.*, a base
rate unaccompanied by other evidence) should not be treated differ-
ently from other cases if one's sole concern is verdict accuracy.[59]
For example, an eighty percent probability of guilt based entirely on
statistical background information involves the same twenty percent
chance of erroneous conviction as an eighty percent probability
based on the testimony of several witnesses.[60]

4. *Effects of Base Rates on Jury Consideration of Other Types of Evidence*

The previous Section argued that base rate evidence is just as
logically probative as case-specific evidence. However, even if the
two are equivalent in principle, one can argue that they have differ-
ent verdict accuracy effects in practice when used by actual, and falli-
ble, fact-finders.

Along these lines, Professor Tribe has argued that jurors will
systematically assign too much weight to base rate evidence. His
claim is that hard statistical data leads decisionmakers to "dwarf the
soft variables" and to assume that "[i]f you can't count it, it doesn't
exist."[61]

---

56   *See* Nesson I, *supra* note 1, at 1196; Nesson II, *supra* note 1.
57   *See* Nesson II, *supra* note 1, at 1391-92.
58   Even L.J. Cohen, one of the harshest critics of the use of probabilistic methods in
legal reasoning, agrees that "relevant" base rate evidence is as probative as case-specific
evidence, since he advocates basing at least some verdicts on frequencies computed for
"suitably narrowed-down" reference classes. *See* L.J. Cohen II, *supra* note 1, at 633.
59   Again, a case may involve more than naked statistical evidence if the absence of
other evidence gives rise to an inference of strategic behavior by one or both parties.
We discuss the problem of strategic behavior further in Section IV. *See infra* notes 96-
103 and accompanying text.
60   The distinction between background and case-specific evidence may be illusory
altogether. Background data may be probative with respect to either prior probabilities
or likelihood ratios, even though the latter are ordinarily thought to reflect case-specific
evidence. *See* Richard Lempert, *The New Evidence Scholarship: Analyzing the Process of Proof*,
66 B.U.L. REV. 439, 454 n.44 (1986) (a mathematical analysis of the gatecrasher
paradox).
61   Tribe, *supra* note 1, at 1361. Oddly, while Tribe expects jurors to attach too

If Tribe's behavioral assumption is correct, then juries might tend to equate particular base rates with the probability of guilt or liability even when other evidence suggests that it would be incorrect to do so. This could substantially weaken the grounds for expecting the admission of base rate evidence to improve verdict accuracy.

However, recent psychological research does not support Tribe's assumption. It suggests that, in a wide range of situations, people generally undervalue base rate evidence and attach too much weight to case-specific evidence.[62]

One explanation for this phenomenon is that people pay more attention to case-specific evidence because it is more anecdotal, personal, and vivid than a set of summary statistics.[63] Others argue that the use of a "representativeness heuristic" explains the effect.[64] Regardless of the sufficiency of these explanations, it appears clear that base rate evidence, rather than case-specific evidence, tends to be "dwarfed" or ignored in cases where both are present.[65]

B.   The Use of Probability Theory to Combine Separate Items of Base Rate Evidence

The previous Section argued that base rate evidence is no less

---

much weight to base rate evidence, he expects the public to attach too little weight to it, and thus to distrust verdicts that overtly reflect its use. *Id.* at 1376. As James Brook notes, he fails to explain this apparent contradiction. *See* Brook 1, *supra* note 1, at 103.

[62]   This tendency to undervalue base rate information has been demonstrated not only in numerous experimental investigations with naive subjects, *see supra* note 29, but also with regard to practicing scientists and physicians. *See, e.g.,* John I. Balla, Robert Iansek & Arthur Elstein, *Bayesian Diagnosis in Presence of Pre-existing Disease,* LANCET, Feb. 9, 1985, at 326; David M. Eddy, *Probabilistic Reasoning in Clinical Medicine: Problems and Opportunities,* in JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES (Daniel Kahneman, Paul Slovic & Amos Tversky eds. 1982); Paul E. Meehl & Albert Rosen, *Antecedent Probability and the Efficiency of Psychometric Signs, Patterns, or Cutting Scores,* 52 PSYCHOLOGICAL BULL. 194 (1955).

[63]   RICHARD E. NISBETT & LEE ROSS, HUMAN INFERENCE: STRATEGIES AND SHORTCOMINGS OF SOCIAL JUDGMENT (1980); Richard E. Nisbett & Eugene Borgida, *Attribution and the Psychology of Prediction,* 32 J. PERSONALITY & SOC. PSYCHOLOGY 932 (1975); Richard E. Nisbett, Eugene Borgida, Rick Crandall & Harvey Reed, *Popular Induction: Information is Not Necessarily Informative,* in COGNITION AND SOCIAL BEHAVIOR 113 (John S. Carroll & John W. Payne eds. 1976). *But see* Shelley E. Taylor & Suzanne C. Thompson, *Stalking the Elusive "Vividness" Effect,* 89 PSYCHOLOGICAL REV. 155 (1982) (arguing that the explanatory value of vividness in other contexts has been overstated).

[64]   A representativeness explanation might hold that individuating, case-specific information is regarded as more similar to, or representative of, the critical features of the issue under consideration than base rate information. *See* Kahneman & Tversky, *supra* note 28; Saks & Kidd, *supra* note 1, at 132.

[65]   For recent verifications of this problem in simulated trial conditions, see JANE GOODMAN, PROBABILISTIC SCIENTIFIC EVIDENCE: JURORS' INFERENCES (1986); David L. Faigman & A.J. Baglioni, Jr., *Bayes' Theorem in the Trial Process,* 12 LAW & HUM. BEHAV. 1, 3 (1988); William C. Thompson & Edward L. Schumann, *Interpretation of Statistical Evidence in Criminal Trials,* 11 LAW & HUM. BEHAV. 167, 167-83 (1987).

relevant than case-specific evidence, and that both are important from the perspective of verdict accuracy. A question still arises, however, as to how base rate evidence should be combined with other evidence in computing a probability for guilt or liability. The overtly probabilistic form of base rate evidence seems to invite the use of mathematical devices such as Bayes' theorem and the product rule when all requisite probabilities are available. However, some have argued that the use of such devices is inappropriate even when it is feasible—for verdict accuracy as well as policy reasons.[66]

Commentators such as Brilmayer, Kornhauser, and L.J. Cohen argue that probabilistic logic is too fundamentally flawed to improve verdict accuracy even when applied to base rates. Instead, they favor what they term "intuitive" decisionmaking strategies that reflect how statistically untrained people actually make decisions. They argue that "intuition" is internally coherent and preferable to probabilistic logic in the event of conflict.

In part, these commentators make a baseline argument: that intuition should generally be presumed correct, unless and until it is proven inadequate, and that probability theory is a departure that requires strong justification. Thus, for example, they denounce the Tversky and Kahneman study of people's responses to the cab example for "arrogantly" assuming that Bayes' theorem provides the correct answer.[67] The following Section discusses the content of the intuitive approach that some anti-Bayesians regard as the appropriate baseline.

1. *"Intuition" as an Alternative to Probabilistic Logic*

One cannot endorse intuition as a baseline approach without raising the question of what it actually is. In general, the meaning that Brilmayer, Kornhauser, and L.J. Cohen assigu to intuition appears to have at least three components: 1) value choices, 2) hasty thinking or heuristic strategies, and 3) empirical beliefs.

a. *Intuitive Value Choices*

Consider again the gatecrasher paradox, relied upon by L.J. Cohen to establish the incorrectness of probability theory. L.J. Cohen regards the case as paradoxical because, despite the apparent 50.1 percent probability that a randomly selected defendant is liable, "our intuitions of justice revolt against the idea that the plaintiff

---

66   *See, e.g.,* Brilmayer & Kornhauser, *supra* note 1.

67   *Id.* at 147. More specifically, they denounce Finkelstein for accepting the results of the study. *See also* L. Jonathan Cohen, *Can Human Irrationality Be Experimentally Demonstrated?*, 4 BEHAV. & BRAIN SCI. 317, 328-29 (1981).

should be awarded judgment on such grounds."[68] He therefore concludes that a "mathematicist" interpretation of probability is fundamentally flawed.[69]

"Intuition" in this sense is merely a way of describing a value choice. Given the probabilities (and assuming no ground for changing them), L.J. Cohen's intuition of justice plainly reflects policy concerns apart from verdict accuracy. He prefers the 50.1 percent chance of a mistake in holding for the defendant to the 49.9 percent chance of a mistake in holding for the plaintiff—based, apparently, on cost of error considerations.[70]

Yet this value choice does not show probabilistic logic to be flawed or "paradoxical" in relation to verdict accuracy.[71] Even if one's intuition about justice always provides the right answer to a question, it can be correct only about the sum of the moral considerations that one considers relevant. For example, an intuition about the gatecrasher problem is correct, if at all, about the proper outcome in light of *all* relevant concerns (both verdict accuracy and separate policy). Yet one's bottom line moral intuition cannot always be correct about each *particular* value (such as verdict accuracy alone) unless one's values always are consistent with each other.[72]

b. *Intuitive Hasty Thinking and Heuristic Strategies*

In some cases, "intuitive" thinking may actually involve nothing more than careless, hasty, ill-informed reasoning. Consider again the cab hypothetical, in which Tversky and Kahneman assume that Bayes' theorem, rather than the responses of their test subjects, provides the correct answer.

The Cohen-Brilmayer-Kornhauser attack on Tversky and Kahneman for preferring "logic" to "intuition" begs the question of whether Bayes' theorem truly is counter-intuitive—or instead is just not immediately obvious. After all, the theorem can be made intuitively plausible through examples, and can be proven logically (*i.e.*, in an intuitively plausible way) by reasoning from intuitively plausible premises. The same is true of the product rule (which indeed

---

68   L.J. Cohen II, *supra* note 1, at 627. Not everyone's intuition is similarly revolted by the prospect of a judgment for the plaintiff. *See, e.g.,* Brook I, *supra* note 1, at 86; Schmalbeck, *supra* note 55, at 222.

69   L.J. COHEN I, *supra* note 1, at 74-81.

70   *See id.* at 75-80; L.J. Cohen II, *supra* note 1, at 632.

71   *See* Brook II, *supra* note 1, at 322.

72   L.J. Cohen rejects the possibility of conflict among the values that are served by trials based on little more than wishful thinking. Thus, he states that holding for the plaintiff when there is a measurable chance of error "hardly seems the right spirit in which to administer justice." L.J. COHEN I, *supra* note 1, at 75. He fails to appreciate that plaintiffs are wronged by incorrect verdicts for defendants and that incorrect verdicts are an inevitable byproduct of uncertainty. *See* Brook II, *supra* note 1, at 318-19.

may be intuitively obvious to begin with). Thus, the failure to apply
either rule may reflect, not a conscious or coherent choice, but in-
stead only a lack of careful thinking—or, more sympathetically, a
heuristic strategy of simplifying information for purposes of making
rapid decisions.[73]

Studies of heuristic strategies reveal that, while these strategies
save time and effort, they sometimes lead to predictably inaccurate
results. Consider, for example, the observation that people often
assess the probability or frequency of an event by the ease with
which relevant instances can be brought to mind.[74] Although this
strategy has intuitive appeal, the relation between frequency and re-
call is imperfect. Research shows that vivid or salient events are
more likely to be recalled than less salient events, even when the
latter are more frequent. Consequently, this strategy may result in
an overestimation of the probabilities associated with vivid or salient
events and an underestimation of the probabilities associated with
less vivid or salient events.

There is no reason to resent logic for demonstrating the weak-
ness of certain approaches that initially are appealing as intuitive
shortcuts. As Saks and Kidd have noted, humanity is hardly de-
graded by its ability to recognize cognitive limitations and to invent
tools that permit an increase in the rigor and power of human
thought processes.[75]

c. *Intuitive Empirical Beliefs*

A final type of intuition relied upon by L.J. Cohen, Brilmayer,
and Kornhauser involves empirical claims that we cannot verify but
consider highly plausible. If probabilistic logic suggests results that
are inconsistent with intuitive empirical beliefs, then one or the
other must be wrong—argnably logic, if one has sufficient confi-
dence in one's intuitive predictions.

The claim of inconsistency with well-founded intuitive empirical
beliefs commonly is expressed in the form of paradoxes that are said
to follow from probabilistic logic. A number of these paradoxes al-
ready have been refuted or shown to be irrelevant to the choice be-
tween "logic" and "intuition."[76] However, we will discuss two

---

73   While some people might refuse to accept probabilistic logic even if it were ex-
plained to them, this refusal might be inconsistent with other beliefs that they hold, and
they might ultimately agree that their beliefs should be internally consistent.

74   *See* Amos Tversky & Daniel Kahneman, *Judgment under Uncertainty: Heuristics and
Biases*, 185 SCIENCE 1124 (1974).

75   *See* Saks & Kidd, *supra* note 1, at 148.

76   *See, e.g.*, Stephen E. Fienberg, *Gatecrashers, Blue Buses, and the Bayesian Representation
of Legal Evidence*, 66 B.U.L. REV. 693 (1986); D.H. Kaye, *Comment: A First Look at "Second-
Order Evidence"*, 66 B.U.L. REV. 701 (1986) [hereinafter *First Look*]; David Kaye, *The Para-*

apparent paradoxes, one purportedly impeaching the product rule and the other, Bayes' theorem, that have not as fully been addressed in the legal literature.

(1) *The Multiple Issue Fallacy*—L.J. Cohen argues that the product rule paradoxically requires accepting the apparently implausible view that plaintiffs almost never are entitled to recover in civil cases that are complex or that involve multiple legal issues.[77] To illustrate his argument, consider a plaintiff who needs to prove ten separate facts in order to be entitled to damages. Assume that there is a ninety percent probability that the plaintiff is correct about each fact. If the facts are independent, then, under the product rule, the posterior probability that the plaintiff should prevail would be $(.9^{10})$, or approximately thirty-five percent. Yet intuitively one might estimate that plaintiffs who are so strongly supported by the facts are more than thirty-five percent likely to be correct.

L.J. Cohen recognizes that the "paradox" does not exist if one can explain the assumed higher percentage by positing that the issues in a multi-faceted cause of action are rarely independent.[78] For example, if in the above hypothetical the plaintiff's ten contentions were entirely dependent, the posterior probability that the plaintiff should prevail would be ninety percent. Cohen's only argument against this solution is that a plaintiff's contentions may be negatively dependent instead of positively dependent (*i.e.*, the truth of each contention may decrease the likelihood of each other contention being true).

While negative dependence no doubt occurs, it is unclear why one should expect it to be more common than positive dependence. Indeed, intuitively one may expect positive dependence to be more common. For example, witnesses who testify for the same party on several issues may have consistent degrees of reliability. Or, the conclusion that a plaintiff believed a contract was formed may increase the likelihood both of contract formation and of plaintiff performance.[79]

---

dox of the Gatecrasher and Other Stories, 1981 ARIZ. ST. L.J. 101; Anne W. Martin, *Comment*, 66 B.U.L. REV. 709 (1986). Some of the asserted paradoxes rest on the claim that probabilistic logic is self-contradictory, rather than inconsistent with intuitive predictions. *See* Kaye, *First Look, supra*, at 705-06.

[77]   L.J. COHEN I, *supra* note 1, at 58-67.

[78]   *See id.* at 61.

[79]   L.J. Cohen supports his argument that negative dependence may be as common as positive dependence with a hypothetical example, involving a plaintiff who claims (1) that a contract was formed, (2) that the plaintiff performed under the contract, and (3) that the defendant did not perform. Cohen argues that evidence supporting contract formation would tend to rebut the claim that the defendant did not perform. Yet there is nothing implausible about inferring plaintiff performance, but not defendant performance, from the fact of contract formation. For example, one could argue that the de-

Thus, the apparent inconsistency between the product rule and intuitive prediction arises only if one makes unrealistic (and indeed, counter-intuitive) assumptions about the typical dependence relationships among a plaintiff's factual claims. One cannot show that the product rule is inconsistent with well-founded intuitive predictions.

(2) *Bayes' Theorem and the Todhunter Paradox*—As demonstrated by the cab example, Bayes' theorem can be somewhat counter-intuitive. Strong case-specific evidence in the form of a high likelihood ratio can give rise to a low posterior probability when combined with a low prior odds ratio. Some advocates of intuitive methods contend that Bayesian combination techniques may, at times, yield posterior probabilities that are not only counter-intuitive, but implausible as well. The so-called Todhunter paradox has been cited by several commentators in support of this view.[80]

The paradox may be described as follows: Suppose that a witness who is correct 99.9 percent of the time claims that the winning number in a lottery of 10,000 tickets was 297. Intuitively, it would seem that we can be 99.9 percent certain that 297 really was the winning number. The number of tickets in the lottery appears to be irrelevant.

However, the Bayesian method requires taking into account the prior probability that 297 would win. This probability clearly does depend on the number of tickets in the lottery, and in this instance is 1 in 10,000. According to L.J. Cohen and others, when this remote prior probability is incorporated into the Bayesian computation, the posterior probability that 297 is the winning number, given that the witness said it was, is very small.[81] Not only is this result counter-intuitive, but experience with announced winning lottery numbers clearly contradicts it as well.

---

fendant intended to accept payment but not to perform, or that performance became disadvantageous to the defendant after the contract was formed due to a change in circumstances. *Id.* at 61-62.

[80] L.J. Cohen, *supra* note 67, at 329; Brilmayer & Kornhauser, *supra* note 1, at 148 n.114; Llkka Niiniluoto, *L.J. Cohen Versus Bayesianism*, 4 BEHAV. & BRAIN SCI. 349, 349 (1981).

[81] According to Bayes' theorem, the posterior probability that the winning ticket is 297 given that the witness says so is computed as follows:

$$P(297/\text{``297''}) = \frac{P(\text{``297''}/297)\ P(297)}{P(\text{``297''}/297)\ P(297) + P(\text{``297''}/\text{not }297)\ P(\text{not }297)}$$

where 297 = the winning ticket is in fact number 297.
      "297" = the witness *says* the winning ticket is number 297.

Under Cohen's proposed Bayesian calculation, the posterior probability indicates only a 9 percent chance that the winning ticket was 297 even though the highly reliable witness says so:

However, the above computation contains a crucial mistake in its application of Bayes' theorem. It confuses (1) the probability that the witness made some mistake (*i.e.*, .1 percent, or 1/1,000) with (2) the probability that he made *this particular* mistake. Yet this second, smaller probability is the one that is of interest here.

What is the probability of this particular mistake, *i.e.*, the probability that the witness says "297" when the winning ticket is in fact some number other than 297? If all possible mistakes were equally likely, then the probability that the witness would say "297" given that he had made a mistake is 1/9,999 (since there are 9,999 possible mistakes in a 10,000 ticket lottery).[82] Therefore, the probability of the witness's particular mistake may be described as P("297"/not 297) = P (mistake) × P("297"/mistake) = 1/1000 × 1/9999.

The correct Bayesian computation for the Todhunter problem suggests a posterior probability of 99.9 percent that the winning ticket was 297.[83] This is well in accord with the intuition and experience that is relied upon to establish the paradox. Thus, the Todhunter problem provides no support for the claim that intuitive predictions are superior to probability theory when the two are inconsistent.[84]

2. *Empirical Grounds for Choosing Between Probabilistic Logic and Intuition*

At present, there is no direct empirical evidence showing whether probabilistic logic or intuition, as applied in the legal setting, leads to greater verdict accuracy. Brilmayer and Kornhauser contend that empirical demonstration in this area is impossible, due to the lack of a truth criterion for trials.[85] That is, since it is rarely, if ever, possible to determine the accuracy of particular verdicts, one

---

$$P(297/\text{``297''}) = \frac{(.999)\,(.0001)}{(.999)\,(.0001) + (.001)\,(.9999)} = .09.$$

[82] In practice, some mistakes may be more likely than others (*e.g.*, an error that contains only a single wrong digit may be more likely than an error in which all three digits are wrong). However, 1/9,999 can be taken as an average because we do not know what the correct ticket is.

[83] The correct Bayesian solution to the Todhunter problem is as follows:

$$P(297/\text{``297''}) = \frac{(.999)\,(.0001)}{(.999)\,(.0001) + (.001)\,(1/9999)\,(.9999)} = .999.$$

For a similar analysis, see Richard D. Friedman, *Route Analysis of Credibility and Hearsay*, 96 YALE L.J. 667, 738 (1987).

[84] Discussions with Joshua Klayman about this problem were especially valuable.

[85] Brilmayer & Kornhauser, *supra* note 1, at 138, 145.

cannot prove that the use of probabilistic logic increases verdict accuracy.

Because of the truth criterion problem, empirical verification of the superiority of probabilistic reasoning over intuitive reasoning in the legal setting is indeed difficult to obtain. However, substantial evidence from other fields where truth criteria exist suggests the superiority of probabilistic methods.[86] Even in decisionmaking areas where only subjective estimates of the relevant probabilities are used, there is some evidence that judgmental accuracy is increased by using Bayes' theorem instead of intuitive methods.[87]

Even though the evidence that Bayesian methods improve judgmental accuracy is taken from areas outside the legal setting, it strengthens the inference that Bayesian methods are superior to intuitive ones and would improve verdict accuracy. At a minimum, the evidence supports shifting the burden of proof to anti-Bayesians to explain why intuitive methods will work better in the law when they yield worse results in all other tested fields.

III

VERDICT ACCURACY AND THE USER OF SUBJECTIVE
PROBABILITIES

In most trials, evidence other than base rates is introduced for consideration by fact-finders. When such evidence is available, or when fact-finders apply their own knowledge and experience to a case, subjective probability issues arise. How credible is an eyewitness identification placing the defendant at the scene of a crime? How likely is it that the defendant made the threats that a witness claims to have heard? How clearly does photographic evidence depict an exchange of money? Even when fact-finders do not make explicit or precise probability estimates about issues, their decisions may be influenced by implicit estimates.

---

86 *See, e.g.,* PAUL E. MEEHL, CLINICAL VERSUS STATISTICAL PREDICTION: A THEORETICAL ANALYSIS AND REVIEW OF THE EVIDENCE (1954); Robyn M. Dawes, *The Robust Beauty of Improper Linear Models in Decision Making,* 34 AM. PSYCHOLOGIST 571 (1979) (graduate admissions); Bruce Duthie & Ken R. Vincent, *Diagnostic Hit Rates of High Point Codes for the Diagnostic Inventory of Personality and Symptoms Using Random Assignment, Base Rates, and Probability Scales,* 42 J. CLINICAL PSYCHOLOGY 612 (1986); Hillel J. Einhorn, *Expert Measurement and Mechanical Combination,* 7 ORGANIZATIONAL BEHAV. & HUM. PERFORMANCE 86 (1972) (patient longevity); Robert Libby, *Man versus Model of Man: Some Conflicting Evidence,* 16 ORGANIZATIONAL BEHAV. & HUM. PERFORMANCE 1 (1976) (business failures); Jack Sawyer, *Measurement And Prediction: Clinical And Statistical,* 66 PSYCHOLOGICAL BULL. 178 (1966) (clinical).

87 *See* Balla, Iansek, & Elstein, *supra* note 62. Admittedly, evidence directly bearing on this claim is scant, although to the extent subjective estimates of the relevant probabilities are in accord with more "objective" estimates, Bayesian methods likely will increase decisional accuracy.

A question arises as to whether rules of probability theory should be applied to combine subjective probability estimates with other information in computing posterior probabilities. Anti-Bayesian commentators have argued that the specification and explicit use of subjective probabilities are inappropriate even when feasible.[88] Although some base their arguments entirely on policy concerns other than verdict accuracy,[89] others do not.[90] This section assesses the use of subjective probabilities in relation to verdict accuracy.

A. The Underlying Debate Concerning the Use of Subjective Probabilities in Statistics

Statisticians have long disagreed about how subjective probabilities, particularly subjective prior probabilities, should enter into decision analyses.[91] To Bayesian statisticians, such probabilities capture informational components of problems, and should therefore be incorporated into formal analyses. Classical statisticians, on the other hand, prefer to restrict their formal analyses to hard data, and to incorporate subjective probability judgments informally prior to making decisions.[92]

The underlying debate between Bayesian and classical statisticians is partly aesthetic and partly practical. The aesthetic issue concerns whether it is appropriate to extend an appearance of mathematical rigor to the analysis of data as vague and personal as subjective probabilities. The practical issue concerns whether formal use of subjective probabilities actually increases judgmental accuracy.

There is no generally correct answer even to the practical issue.[93] The effect of mathematical logic on judgmental accuracy depends on how well the available information correlates with reality. A subjective probability estimate that is relatively well-founded (*e.g.*, based on reliable and extensive observation) may be relatively likely to improve judgmental accuracy. As an estimate becomes less well-founded, its capacity to improve decisional accuracy may decline.

However, neither classical nor Bayesian statisticians view intui-

---

88 The specification and use of subjective probabilities raise practical problems involving computational complexity, and translation difficulty (*i.e.*, meaningfully converting one's judgments about evidence into numerical terms). We address these problems in Section IV, *infra*.

89 *See, e.g.*, Nesson II, *supra* note 1.

90 *See, e.g.*, Brilmayer & Kornhauser, *supra* note 1.

91 *See, e.g.*, Allan Birnbaum, *On the Foundations of Statistical Inference*, 57 J. AM. STATISTICAL A. 269 (1962); Bross, *Discussion of A. Birnbaum's Paper, "On the Foundations of Statistical Inference," id.* at 309.

92 *See, e.g.*, Harry V. Roberts, *"Bayesian Inference,"* in 2 INT'L ENCYCLOPEDIA SOC. SCI. 28 (1968).

93 *But see supra* notes 68-79 and accompanying text.

tive combination methods as superior to probability theory. Classical statisticians argue only that there is little to be gained from using probability theory with respect to subjective probabilities, not that intuition works better.

B. Advantages of Using Probability Theory at Trial with Regard to Subjective Probabilities

Inaccurate information is likely to yield inaccurate verdicts regardless of the aggregation technique employed. In such cases, it is difficult to argue that one method is superior to another in producing accurate outcomes. However, good reasons remain—even apart from the empirical evidence discussed above—for expecting that mathematically rigorous techniques will, in general, yield more accurate verdicts than other methods such as unaided intuition.

This expectation arises from two advantages of using mathematical logic. First, it enables one to be internally consistent and logical in deriving conclusions from one's information and beliefs. Second, it ameliorates difficulties in combining "hard" with "soft" evidence.

1. Internal Consistency

Applications of probability theory such as the product rule and Bayes' theorem are derived from a handful of simple axioms. By providing coherent means for combining separate items of information, they permit decisionmakers to reach conclusions that are consistent with their judgments about all of the available information.

In contrast, intuitive methods provide no guidelines for combining judgments about separate items of information. Lacking guiding axioms, they permit a decisionmaker to reach virtually any conclusion on the basis of any set of evidence.

To be sure, greater internal consistency does not guarantee greater verdict accuracy. An increase in the internal consistency of jury verdicts conceivably could lead to reduced verdict accuracy, if, for example, jurors' beliefs about subsidiary issues of fact are commonly wrong. However, if there is a positive correlation between jury judgments about such issues and actual states of affairs, verdict accuracy is likely to be enhanced by procedures that ensure internal coherence.[94]

---

94 The view that there is no positive correlation between jury judgments about subsidiary issues of fact and actual states of affairs would appear to have implications going far beyond the choice between probability theory and intuition. Absent a positive correlation, verdict accuracy might be so low as to suggest that the trial system is not meeting minimum standards of acceptable dispute resolution.

2. *Combining "Hard" With "Soft" Evidence*

In addition to improving internal consistency, the use of mathematical logic with respect to subjective probabilities may ameliorate the difficulties that fact-finders experience in combining "hard" with "soft" evidence. Whether one agrees with Tribe's intuition about the tendency of fact-finders to "dwarf the soft variables," or with the empirical research suggesting that it is the hard data that get slighted, the use of mathematical logic can assure that neither type of evidence dwarfs the other.[95] No such assurance is provided by alternative combination methods.

IV

PROBLEMS WITH THE USE OF OVERTLY PROBABILISTIC
EVIDENCE AND METHODS

Although the use of overtly probabilistic evidence and methodologies generally can improve verdict accuracy, in some situations their benefit is more questionable. This section describes some of the problems associated with probabilistic evidence and procedures that restrict their usefulness in the courtroom.

A.   Complexity and Translation Difficulties

Previous commentators have noted the unmanageable complexity that can result from attempting to apply rules of mathematical logic to multiple factual issues, beliefs about the issues, and underlying items of evidence.[96] Where such complexity is great enough, the likelihood of error or misapplication of the rules may be large enough to offset any verdict accuracy benefit (even disregarding separate policy concerns such as attempting to minimize the length and cost of trials).

Computational complexity is only one aspect of what may often be the prohibitive difficulty of applying rules of mathematical logic in the courtroom. A more fundamental problem is that fact-finders are likely to experience difficulty in translating all relevant evidence into numerical terms. Since few people are accustomed to statistical ways of thinking and reasoning, there is a danger that one's opinions and judgments may be altered when they are restated numerically.

A related problem is that some of the probabilities needed to

---

95   It is somewhat ironic that those who oppose the use of formal combination techniques such as Bayes' theorem frequently lament the difficulties that juries have combining the various types of hard and soft data that are presented at trial.

96   *See, e.g.,* Craig R. Callen, *Notes on a Grand Illusion: Some Limits on the Use of Bayesian Theory in Evidence Law,* 57 IND. L.J. 1 (1982); Tribe, *supra* note 1.

perform particular mathematical operations may be unavailable or difficult to specify. For example, several probabilities are required to compute posterior probabilities under Bayes' theorem,[97] and, further, even if base rate evidence is available for both it may be unclear *which* base rates should be used.[98] Furthermore, the conditional probabilities that compose the Bayesian likelihood ratio may be difficult to conceptualize, let alone quantify.

For example, in assessing the likelihood that a defendant committed a crime given a set of evidence, a Bayesian fact-finder must be able to specify both the prior probability that the defendant committed the crime and the likelihood that the set of evidence would be present both if he did and if he did not commit the crime. Not only are such probabilities hard to think about and specify, but there is evidence that people fail to understand their relevance for posterior probability computations even when their values are supplied.[99]

Further difficulties arise in attempting to specify the degree of dependence between two events in order to apply the product rule. For example, in a contract case where the only contested issues are contract formation and plaintiff performance, even jurors who can specify a subjective probability for each issue may have difficulty specifying how plaintiff correctness on the former issue affects the likelihood of plaintiff correctness on the latter.

Translation difficulties suggest that mathematical tools will not assist fact-finders in all (or perhaps even many) cases. These difficulties often may be severe enough to offset any expected benefit to verdict accuracy from applying logical techniques.

However, this does not weaken the expectation of benefit in cases where the difficulties are relatively slight. Consider cases where there are only a few items of evidence, one of which is overtly probabilistic and relates, for example, to paternity or to the identification of blood, hair, and skin samples found at the scene of an incident.[100] In addition, consider cases where one could reasonably argue that the elements of the plaintiff's case are independent. Instruction about the product rule may prevent juries from holding

---

97   In order to compute P(H/D), one needs to know P(H), P(not H), P(D/H), and P(D/not H). *See supra* note 84. However, since P(H) + P(−H) = 1, one need only know one prior probability to compute the other. Furthermore, if one is presented with a single value for the likelihood ratio, one need not be concerned with the values of the individual conditional probabilities P(D/H) and P(D/−H) that compose it.

98   *See supra* text accompanying notes 30-87.

99   *See* Ruth Beyth-Marom & Baruch Fischhoff, *Diagnosticity and Pseudodiagnosticity*, 45 J. PERSONALITY & SOC. PSYCHOLOGY 1185 (1983); Michal E. Doherty, Clifford R. Mynatt, Ryan D. Tweney & Miael D. Schiavo, *Pseudodiagnosticity*, 43 ACTA PSYCHOLOGICA 111 (1979).

100   Such evidence may provide, for example, the percentage of the population that, like the defendant, would be a possible source of the tested genetic material.

for a plaintiff who is likely to be correct about each element of his or her claims, but who is not likely to be correct overall.[101]

Yet, even in the great majority of cases where juries cannot be expected to make rigorous mathematical computations, instruction about probability theory (either by an expert or from the bench) may be helpful. Such instruction may permit juries to understand the significance of base rate evidence, and more generally to combine "hard" and "soft" evidence without unduly favoring either. Thus, such instruction may improve verdict accuracy even if used relatively loosely as a guide to intuitive judgments.

Judges also may benefit from understanding probability theory. For example, such understanding might improve decisions about the admissibility of base rate evidence where the only consideration is balancing its probative value against undue prejudice or delay. Further, when a knowledgeable judge rather than a jury is the fact-finder, more extensive and formal reliance on probability theory may be feasible.

B. Feedback Problems with the Use of Base Rates

In Section II, we argued that verdict accuracy generally can be improved by incorporating base rate evidence into the trial process. However, the use of such evidence can give rise to feedback, or opportunistic responses to the knowledge that such evidence is being used. As an example, consider permitting prosecutors to introduce base rate evidence showing the percentage of defendants, in trials concerning similar crimes, who were actually guilty (assuming that this could somehow be determined). Admitting this evidence in order to suggest a prior probability of guilt, however inconsistent with the constitutionally based presumption of innocence, might increase verdict accuracy.

However, it may be that one reason for the frequency of actual guilt among such defendants is that prosecutors, in light of the presumption of innocence, do not indict unless the evidence of guilt is relatively strong. If prosecutors subsequently began indicting based on weaker evidence, then the base rate frequency might change. In other words, the very fact that the base rate evidence was being used could cause this frequency to change and diminish its probative value.[102]

---

[101]   Jury instructions seem to require juries to look at each element separately, not at the aggregate of the plaintiff's case. Nesson II, *supra* note 1, at 1386-88, defends such instructions on policy grounds.

[102]   Admittedly, the base rate frequency could be continually updated to ensure that jurors receive accurate information. Yet, there might be accuracy concerns about the updating.

Feedback effects are difficult to factor into decision analyses because one often has no way of knowing their direction or magnitude. Indeed, it is hard to predict when such effects are most likely to occur (with regard to both base rates and other kinds of evidence). The possibility of such effects reminds us, however, that the admissibility of base rate evidence is not a straightforward matter, from either a verdict accuracy or a general policy perspective.[103]

V
SUMMARY AND CONCLUSIONS

This Article examined the capacity of overtly probabilistic evidence and methodologies to improve verdict accuracy. Policy and other practical reasons for avoiding an overtly probabilistic trial process have been downplayed in order to explore the idea that base rates and other probabilistic evidence are rich sources of information that, at present, are not always understood in legal circles.

Base rate evidence, while neither a panacea nor devoid of problems, is relevant to the truth of asserted facts, and indeed is no less relevant in principle than case-specific evidence. An eighty percent probability of guilt based entirely on base rate information carries with it the same twenty percent chance of a false conviction as an eighty percent probability of guilt based on, say, the somewhat unreliable testimony of an eyewitness. All evidence contains a risk of error if relied upon to support a conclusion; overtly probabilistic evidence simply makes the risk more visible.

A wide range of base rates are potentially relevant to a question of fact. Some base rates are more informative or reliable than others because they are based on more appropriate reference classes. However, from a verdict accuracy standpoint, a base rate need not meet any minimum standard of specificity in order to have evidentiary weight. There usually is no such thing as *the* correct base rate for a particular issue. Even base rates that are relatively unspecific, or that are based on relatively small samples, may provide the best estimate for a given frequency in the absence of other information.

Overtly probabilistic methodologies that employ rules of probability theory (such as the product rule or Bayes' theorem) also can improve verdict accuracy. In combining multiple items of evidence, intuitive strategies are often inconsistent with and inferior to mathematical techniques. Even with regard to subjective probabili-

---

103   The strategic introduction or withholding of information is a concern with *all* types of evidence presented at trial. Whether or not the consequences of such strategizing are most serious when statistical evidence is involved is an empirical question that, to our knowledge, has not been tested.

ties, probability theory is likely to improve verdict accuracy when there is a positive correlation between jury probability estimates and actual states of affairs.

To be sure, there are situations where overtly probabilistic methodologies may bring little benefit. Sometimes fact-finders may have difficulty using them correctly due to complexity and translation difficulties. The use of base rates may also cause feedback problems.

However, there appears to be little evidence that base rate information will "dwarf the soft variables," leading jurors to pay too little attention to case-specific evidence. Psychological research suggests instead that base rate evidence is more likely to be underweighted. Yet whether jurors pay too little attention to the "hard" evidence or the "soft," the difficulties which they experience in combining the two provide a strong reason for acquainting them with mathematical rules that treat the two as equivalent.

In closing, we wish to reiterate that we do not claim that overtly probabilistic evidence and methods should be incorporated into trials. Nor do we claim that policy concerns other than verdict accuracy are unimportant. However, we do claim that the use of an overtly probabilistic approach is neither logically nor theoretically inappropriate.

Commentators who oppose such an approach even when it is feasible are urging a sacrifice of verdict accuracy in order to advance other policy objectives. There is nothing wrong in principle with such a tradeoff, and it may be appropriate in many instances. But this tradeoff requires explicit acknowledgement and justification based on a thorough understanding of the benefits of overtly probabilistic evidence and methods.