

Michigan Law Review

Volume 75
Issue 5 *Issues 5&6*

1977

Modeling Relevance

Richard O. Lempert
University of Michigan Law School, rlempert@umich.edu

Follow this and additional works at: <https://repository.law.umich.edu/mlr>



Part of the [Courts Commons](#), and the [Evidence Commons](#)

Recommended Citation

Richard O. Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021 (1977).
Available at: <https://repository.law.umich.edu/mlr/vol75/iss5/11>

This Essay is brought to you for free and open access by the Michigan Law Review at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in Michigan Law Review by an authorized editor of University of Michigan Law School Scholarship Repository. For more information, please contact mlaw.repository@umich.edu.

MODELING RELEVANCE

Richard O. Lempert*†

During the past decade, particularly during the years immediately following the California Supreme Court's decision in *People v. Collins*,¹ a number of articles have appeared suggesting ways in which jurors might use certain mathematical techniques of decision theory as aids in the rational evaluation of circumstantial evidence.² Professor Tribe, in an important response to the post-*Collins* articles, argues against introducing these techniques into the factfinding process. Problems that Tribe foresees include the necessary imprecision of the probabilistic estimates that these techniques require, the dwarfing of soft variables by those that are more readily quantified, and the potential dehumanization of the trial in the name of rational factfinding.³

I find Tribe's arguments convincing: with certain narrow exceptions "the costs of attempting to integrate mathematics into the factfinding process of a legal trial outweigh the benefits."⁴ This judgment is apparently shared by others, for the spate of articles following *Collins* diminished substantially after Tribe's published response.⁵ However, mathematics relates to trial processes in a way that Tribe's article does not address: mathematical models may serve as heuristic devices.⁶ As a language, mathematics can help

© Richard O. Lempert 1977.

* Professor of Law, The University of Michigan. A.B. 1964, Oberlin College; J.D. 1968, Ph.D. 1971, The University of Michigan.—Ed.

† I would like to thank Don Regan, Mel Guyer, and Judith Lachman for their careful reading of this paper and their many helpful suggestions. Work on this article was supported in part by the Cook Funds of The University of Michigan Law School.

1. 68 Cal. 2d 319, 438 P.2d 33, 66 Cal. Rptr. 497 (1968).

2. See, e.g., the articles cited in Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1332 n.5 (1971). For those interested, the Tribe article contains a description of the *Collins* case. *Id.* at 1334-37.

3. *Id.* at 1393.

4. *Id.* at 1377.

5. Some who have continued writing have become more sensitive to the problems involved in offering mathematical arguments as an aid to jury decision making and more cautious in what they advocate. See, e.g., Fairley, *Probabilistic Analysis of Identification Evidence*, 2 J. LEGAL STUD. 493 (1973).

6. Tribe recognizes this possibility and reserves the right to object to it. See Tribe, *supra* note 2, at 1331 & n.4.

clarify those legal rules that involve weighing evidence in an essentially probabilistic fashion.⁷

In this article I try to show the utility of two simple models, Bayes' Theorem and regret matrices, for thinking about the meaning of relevance and for analyzing those evidentiary rules, which I call the "relevance rules," generally associated with this topic.⁸ The discussion assumes that the factfinder is a jury and, unless otherwise noted, that the issue to be resolved is a defendant's guilt. However, the analysis may be readily generalized to the situation where the factfinder is a judge and/or a question other than guilt is at issue. The first section of this article applies the two models to a simplified situation where the factfinder must evaluate only one item of indisputably accurate testimony.⁹ The second section explores complexities that can arise when a case involves two or more items of possibly unreliable evidence.

I. MATHEMATICAL MODELS AND THE RELEVANCE RULES

A. Bayes' Theorem

First we must attend to Bayes' Theorem. This theorem follows directly from two elementary formulas of probability theory: if A and B are any two propositions, then:

$$P(A \& B) = P(A|B) \cdot P(B)^{10} \quad (1)$$

$$P(A) = P(A \& B) + P(A \& \text{not-}B)^{11} \quad (2)$$

7. Daniel Kornstein, for example, has used Bayes' Theorem as an aid in analyzing the problem of harmless error. Kornstein, *A Bayesian Model of Harmless Error*, 5 J. LEGAL STUD. 121 (1976). A second model used in this article, the regret matrix, a form of utility matrix, is also helpful in thinking about the problem of harmless error. See also Ball, *The Moment of Truth: Probability Theory and Standards of Truth*, 14 VAND. L. REV. 807 (1961); Kaplan, *Decision Theory and the Factfinding Process*, 20 STAN. L. REV. 1065 (1968).

8. See R. LEMPERT & S. SALTZBURG, *A MODERN APPROACH TO EVIDENCE* 148-53 (1977). Portions of this article reproduce arguments made there. By "relevance rules" I mean rules like those codified under Article IV of the Federal Rules of Evidence. They relate to evidence of character, habit, subsequent remedial measures, compromises and offers to compromise, payment of medical expenses, pleas and offers to plea, and liability insurance.

9. In the first section of this paper I shall generally ignore problems that exist because evidence is always received in a context that includes other evidence. The discussion shall proceed as if the evidence in question were the last piece of evidence received in a trial and as if the probability of receiving that evidence were conditionally independent of all the evidence previously received. In addition, I shall assume that the evidence discussed presents no problems of veracity or authenticity. These assumptions simplify the discussion in the text.

10. These symbols mean that the probability that events A and B will both occur is equal to the probability that A will occur if B has occurred times the probability that B will occur. For example, if A = a warm day and B = a sunny day, the probability that it will be both warm and sunny equals the probability that it will be warm if it is in fact sunny times the probability that it will be sunny.

11. These symbols mean that the probability that an event A will occur equals

From these rather basic equations the following formula may be derived:

$$O(G|E) = \frac{P(E|G)}{P(E|\text{not-G})} \cdot O(G)^{12} \quad (3)$$

This formula describes the way knowledge of a new item of evidence (E) would influence a completely rational decision maker's evaluation of the odds that a defendant is guilty (G).¹³ Since the law assumes that a factfinder *should* be rational, this is a normative model; that is, the Bayesian equation describes the way the law's ideal juror evaluates new items of evidence.¹⁴ What this equation says is that the odds (O) that a defendant is guilty, given the introduction of a new item of evidence, is equal to (1) the probability that the evidence would be presented to the jury if the defendant is in fact guilty, (2) divided by the probability that that evidence would be presented

the probability that event A will occur with event B plus the probability that event A will occur with any event that is not B. If A = a warm day and B = a sunny day, the probability that it will be a warm day equals the probability that it will be warm and sunny plus the probability that it will be warm and not sunny.

12. Bayes' Theorem follows directly from the equations given in the text at notes 10 & 11 *supra*. Expressing these in terms changed from A's and B's to G's (guilt) and E's (new evidence) to fit the paradigm case, a criminal trial in which the issue is the defendant's guilt:

$$(1) P(G \& E) = P(G|E) \cdot P(E)$$

$$(2) P(G) = P(G \& E) + P(G \& \text{not-E})$$

it can be shown that

$$(3) P(G|E) = \frac{P(E|G)}{P(E)} \cdot P(G)$$

and

$$(4) P(E) = P(E|G) \cdot P(G) + P(E|\text{not-G}) \cdot P(\text{not-G}).$$

Using (4) to calculate P(E) in (3), we obtain

$$(5) P(G|E) = \frac{P(E|G)}{P(E|G) \cdot P(G) + P(E|\text{not-G}) \cdot P(\text{not-G})} \cdot P(G).$$

Equation (5) is one form of Bayes' Theorem. If O(G) represents the "odds of G," defined as P(G)/P(not-G), then (5) can be rewritten as

$$(6) O(G|E) = \frac{P(E|G)}{P(E|\text{not-G})} \cdot O(G),$$

the form of the theorem that appears in the text.

13. The symbol G could as easily be L for liable, N for negligent, or M for any matter in issue.

14. One might define two normative models of jury behavior. From the standpoint of the law of evidence the normative model implicit in most discussions of jury factfinding is the model of the "rational juror" described in the text. From the standpoint of the legal system one might argue that we employ jurors in large part because we want to inject values other than rationality into the factfinding process. See generally H. KALVEN & H. ZEISEL, *THE AMERICAN JURY* (1966).

to the jury if the defendant is in fact not guilty,¹⁵ (3) times the prior odds¹⁶ on the defendant's guilt. The prior odds are the odds that would have been given of the defendant's guilt before receipt of the item of evidence in question.

For example, suppose at some point in a criminal trial the factfinder believes that the odds are fifty-fifty, or 1:1, that the defendant is guilty. A more familiar way of stating this is that the factfinder believes that the probability of the defendant's guilt is .50.¹⁷ The evidence next received proves the following: that the perpetrator's blood, shed at the scene of the crime, was type A; that the defendant's blood is type A; and that fifty per cent of the suspect population¹⁸ has type A blood. Thus, if the defendant were the perpetrator the probability that the blood found at the scene would be type A is 1.0.¹⁹ The probability that the blood would be type A if someone else committed the crime is .50, or ½, since half of the other possible suspects have type A blood. Plugging these figures into the formula indicates that after receiving the evidence on the blood a rational decision maker would evaluate the odds of guilt as:

$$O(G|E) = \frac{1}{.5} \cdot \frac{1}{1} = \frac{1}{.5} = 2:1.$$

15. In this section it is assumed implicitly that the probability that evidence would be presented to the jury is the same as the probability that the evidence exists.

16. The figure for these odds is not important to the following analysis, though it might be very important in analyzing other problems such as harmless error. It seems unlikely that jurors consciously think in terms of the odds of guilt after each item of evidence is received. Yet it may well be that, without stopping to quantify, they are influenced to make incremental changes in their perception of the parties' chances after hearing items of evidence in much the way Bayes' Theorem suggests, and, if asked, they may be able to express these odds in mathematical terms. See, e.g., Weld & Roff, *A Study in the Formation of Opinion Based upon Legal Evidence*, 51 AM. J. PSYCH. 609 (1938). Cf. J. THIBAUT & L. WALKER, *PROCEDURAL JUSTICE* chs. 6 & 7 (1975).

17. Many find Bayes' Theorem more intuitively understandable when expressed in terms of probabilities, as in equation (5) in note 12 *supra*, than when it is expressed in terms of odds.

18. The suspect populations could be people in the United States, people in a particular locality, males in a locality, black people, white people, etc., based upon what already has been proved about the characteristics of the perpetrator. The textual example assumes that the suspect population is relatively large.

19. At this point some might object that it can never be completely clear that the blood found was the perpetrator's. The point is well taken and is a reason why (1) I don't advocate using this model as an aid to jury factfinding, and (2) I have stipulated that the matter has been proved. The fact that absolute certainty may never exist with respect to an item of evidence does not affect the basic argument of this section. Certain implications of this fact will be discussed in the following section when I introduce the idea of conditional independence and talk about cases involving several items of evidence.

The new evidence has raised the odds in favor of the defendant's guilt to 2:1. Another way of stating this result is that the factfinder's best estimate of the probability that the defendant is guilty is now .67. Evidence that changes an estimated probability of guilt in this fashion is clearly relevant in a criminal trial.

Consider another case. Assume that the range of possible suspects has been limited to voters in a community so conservative that only one out of ten voters supports the liberal candidate. While a group of conservative jurors drawn from this community might be angered by evidence that the defendant supports the liberal candidate, such a showing would not influence the judgment of an ideal juror. Absent some reason to believe that liberals are more prone to commit the crime in question, the probability that the defendant could have been shown to be a liberal were he guilty is .1, the same as the probability that he could have been shown to be a liberal were he not guilty. Solving the Bayesian equation we find:

$$O(G|E) = \frac{.1}{.1} O(G) = O(G)$$

The odds on the defendant's guilt remains $O(G)$; the same as they were before the jury learned of the defendant's political affiliation. In these circumstances evidence of the defendant's political affiliation is not relevant.

1. *Logical Relevance*

In both examples the effect of the evidence on the decision maker's final judgment as to guilt turns entirely on the ratio $\frac{P(E|G)}{P(E|\text{not-G})}$, conventionally called the *likelihood ratio*. In the first example $P(E|G)$ was twice $P(E|\text{not-G})$, and the factfinder doubled his prior odds of the defendant's guilt. In the second example $P(E|G)$ and $P(E|\text{not-G})$ were the same, so the likelihood ratio was one and the factfinder's prior estimate of the defendant's guilt remained unchanged. In terms of the Bayesian model, it will always be the case that the impact of new evidence on prior odds on guilt, or on any other disputed hypothesis, will be solely a function of the likelihood ratio for that evidence. Where the likelihood ratio for an item of evidence differs from one, that evidence is *logically relevant*. This is the mathematical equivalent of the statement in *Federal Rules of Evidence* (FRE) 401 that "relevant evidence" is "evidence having any tendency to make the existence of any fact

that is of consequence to the determination of the action more probable or less probable than it would be without the evidence."²⁰ Hence, evidence is logically relevant only when the probability of finding that evidence given the truth of some hypothesis at issue in the case differs from the probability of finding the same evidence given the falsity of the hypothesis at issue. In a criminal trial, if a particular item of evidence is as likely to be found if the defendant is guilty as it is if he is innocent, the evidence is logically irrelevant on the issue of the defendant's guilt.

As a practical matter courts may be justified in rejecting evidence as logically irrelevant when the likelihood ratio is only *slightly* different from one, since such evidence will have little effect on the odds that the disputed hypothesis is true.²¹ A slight difference in this context must be very small indeed, since a likelihood ratio of 1.5 would lead a factfinder to increase by fifty per cent the estimate of the odds in question and a likelihood ratio of 2.0 would result in a doubling of the prior odds.²²

It is clear from the model that the likelihood ratio depends entirely on the relative magnitudes of $P(E|G)$ and $P(E|\text{not-}G)$ and not on the absolute magnitude of either. Thus evidence that is very unlikely to be associated with a guilty defendant will nevertheless be probative of guilt so long as the evidence is more (or less) likely to be associated with an individual who is not guilty. Suppose, for example, that in an assault case it can be shown both that the defendant is a heroin addict and that one out of 500 criminal assailants are heroin addicts. The latter information means that it is very unlikely that any given criminal assailant is a heroin addict. However, if it can also be shown that of the people who never engage in criminal

20. FED. R. EVID. 401 (emphasis added).

21. A court will often be unable to specify the precise likelihood ratio that is appropriate given the evidence and the issue in question. What a judge may be able to sense is that, although the likelihood ratio may take on any of a range of values, the most probable value of the ratio is one and that it would be unreasonable for a jury to find the likelihood ratio to be more than slightly different from one. When this is the case, the court is justified in excluding the evidence on the ground of logical irrelevance. When a reasonable jury could find the appropriate likelihood ratio to be more than slightly different from one, the jury's responsibility for weighing evidence precludes the court from excluding the evidence as logically irrelevant even if the court believes that the most probable likelihood ratio is one or very close to it.

22. Since we are assuming the evidence is the last evidence received, the argument clearly holds except where the prior odds on guilt are at the very threshold of being sufficient to convict. If additional evidence was still to be offered it is possible that a number of items of evidence that were individually of low relevance would, when taken together, be of considerable relevance.

assault only one in 1000 are heroin addicts, knowledge that the defendant is an addict should result in a doubling of the prior odds that the defendant was the assailant. Conversely, if it could be proved that for every 250 nonassailants there is one heroin addict, evidence of the defendant's addiction and the rate of criminal assault among addicts should lead to a halving of the prior odds that the defendant is guilty of assault. In either of these supposed cases there may be good reason to keep evidence of the defendant's addiction from the jury, but the reason is not that the information standing alone is logically irrelevant.²³

2. Estimation Problems

Courts declare evidence irrelevant for several reasons. Sometimes they are concerned that the likelihood ratio may be one or very close to it. This problem, examined above, is properly called the problem of "logical relevance." On other occasions courts are concerned with the possibility that the factfinder will misestimate the probabilities that make up the likelihood ratio; *i.e.*, $P(E|G)$ and/or $P(E|\text{not-}G)$. Overestimating the numerator or underestimating the denominator makes the conclusion sought by the proponent of the evidence appear more probable than it actually is; underestimating the numerator or overestimating the denominator has the opposite result. In the assault hypothetical presented above, if the factfinder thought that the probability that a nonassailant would be a heroin addict was one in 10,000 rather than one in 1,000, this misestimation would lead to a twentyfold increase in the odds that the defendant was the assailant rather than the twofold increase that was in fact justified. I call such problems "estimation problems."

Estimation problems take several forms. The most obvious is that evidence may be given more weight than it deserves. The jurors may exaggerate the probative value of the evidence because they believe that the association between evidence and hypothesis is more powerful than it in fact is or because they are not estimating probative worth in the context that is proper given the facts of the case. The heroin example of the preceding paragraph is a situation in which the jurors misestimate the strength of a crucial association, throwing the denominator of the likelihood ratio off by a factor of ten. FRE 404, the general rule excluding character evidence, is a relevance rule that can be justified, in part, on this ground. When

23. The evidence is likely to be prejudicial. Also, in the context of other evidence it may be irrelevant for reasons discussed in section II of this article.

courts reject evidence because of this type of estimation problem, they often categorize the problem as one of prejudice, a term I prefer to reserve for another situation,²⁴ or they may speak of the danger of confusing or misleading the jury.

Courts rarely recognize explicitly the danger that jurors will misestimate the probative value of evidence by failing to appreciate the context in which the evidence should be evaluated.²⁵ However, several of the relevance rules may be justified, in part, because the evidence they exclude appears likely to raise such problems.²⁶ For example, rules like FRE 410, excluding evidence of withdrawn guilty pleas, are often justified on the ground that the excluded evidence *is too probative*: to admit evidence of a guilty plea after allowing the plea to be withdrawn would effectively cancel the benefits of the right to withdraw.²⁷ The presumed probative value of the plea is used to justify decisions admitting this evidence where withdrawal does not depend upon a showing that the plea was coerced or otherwise improperly elicited and is an argument against the federal rule of general exclusion. Attention to estimation problems suggests, however, that the federal approach has much to commend it. Jurors might well perceive the numerator of the likelihood ratio for this evidence as the probability that a guilty person would plead guilty and the denominator as the probability that an innocent person would plead guilty. The first, given known rates of guilty pleas, might be reasonably estimated by jurors to be anywhere between .1 and .9. The second would probably be given a very low value, .001, for example. Dividing numerator by denominator suggests that the evidence is quite probative. But these hypothetical jurors would in fact be estimating the likelihood ratio for only part of the evidence before them. In the context of the case, the probabilities that the jurors should be evaluating are the probability that a plea of guilty would

24. See text at notes 36-53 *infra*.

25. This judicial attitude is often defensible, for ordinarily it is the task of the opposing counsel to put a party's evidence into context. However, when the jurors' intuitions are likely to be grossly inaccurate, setting evidence in context may require substantial attention to collateral matters. In these circumstances the decision to exclude evidence rather than open up collateral issues may be justified.

26. Elsewhere I argue that a feature characterizing most of the relevance rules is that they can be justified on several different grounds. Thus the likelihood of estimation problems differs for each of the relevance rules, and no rule is justified *solely* on the ground that the evidence it excludes is likely to pose estimation problems. For a further exposition of these views, see R. LEMPERS & S. SALTZBURG, *supra* note 8, ch. 4.

27. Cf. MCCORMICK'S HANDBOOK OF THE LAW OF EVIDENCE 635 (2d ed. E. Cleary 1972) [hereinafter cited as MCCORMICK].

be made *and later withdrawn* if the defendant were in fact guilty and the probability of the same event if the defendant were in fact innocent. The first probability is likely to be *substantially* less than the probability that a guilty person would plead guilty. The second probability, while necessarily less than the probability that an innocent person would plead guilty, would not be as substantially decreased.²⁸ Jurors who fail to appreciate the information conveyed by the fact that a plea of guilty is withdrawn will give withdrawn guilty pleas considerably more weight than they deserve.²⁹ I believe the likelihood that jurors will estimate the wrong set of probabilities is sufficiently great as to be one justification for rules like FRE 410.³⁰

An estimation problem also exists when there is so little information about the relationship of certain evidence to the hypothesis in question that the implications of the evidence are unclear. In these circumstances courts often exclude evidence as irrelevant rather than let the jurors speculate on its import. Since such evidence might well relate to the probability of guilt or innocence if its true implications were known, a more precise justification for exclusion is "relevance unknown." If the textual example that posited a relationship between heroin addiction and assault did not ring true, it is probably because we lack the base rate information needed to evaluate the relationship between heroin addiction and the likelihood of engaging in an assault. Although the image of the "dope fiend"

28. This probability is necessarily less because it is a probability of two events: (A) a plea of guilty would be made, and (B) the plea would be later withdrawn. $P(A \& B)$ can never be greater than $P(B)$. What is crucial, however, is the *ratio* between the probability that a plea would be entered and later withdrawn if the defendant were guilty and the probability of the same event if the defendant were innocent. If the empirical assumptions that underlie the textual argument are correct, this ratio will be much closer to 1:1 than the ratio of the probability that the defendant would plead guilty if he were guilty to the probability that the defendant would plead guilty if he were innocent.

29. The argument in the text conceptualizes the two aspects of the evidence, that the plea was made and that it was later withdrawn, as if they are aspects of one item of evidence. This appears reasonable because the jurors are likely to receive the information as if it were a single fact and, I believe, are likely to treat it as such. However, one could also conceptualize this as a situation involving two discrete items of evidence, one being that a plea was made and the other that it was later withdrawn. When the evidence is conceptualized in this way, the approach taken in section II applies.

30. For those who accept my analysis of the probabilities involved but believe that I underestimate the perceptiveness of jurors, let me point out that a number of intelligent commentators have apparently made the mistake I expect of jurors. McCormick, for example, writes, "[I]t may be argued, a plea of guilty if freely and understandingly made is so likely to be true that to withhold it from the jury seems to ask them to do justice without knowledge of one of the most significant of the relevant facts." MCCORMICK, *supra* note 27, at 635.

is that of a violent personality, effects associated with addiction suggest that addicts are less likely than nonaddicts to engage in physical violence for its own sake.³¹ With no good evidence of appropriate base rates and conflicting images of the violent propensities of heroin addicts, it makes sense to keep evidence of heroin addiction from a jury in assault cases because its relevance is unknown.³²

Under FRE 403 and at common law, courts have discretion to exclude logically relevant evidence likely to pose estimation problems if the probative value of the evidence is substantially outweighed by the danger that it will mislead the jury. The Bayesian model suggests that in exercising this discretion the more the court's estimate of the proper likelihood ratio for an item of evidence deviates from 1:1 the less willing the court should be to exclude that evidence. If the likelihood ratio for an item of evidence is 2:1 and the factfinder perceives it as 20:1 the misevaluation might well be of critical importance. However, if the likelihood ratio for the evidence is 100:1 and the factfinder misperceives it as 1000:1, the error is less likely to be critical because the evidence whether properly weighed or overweighed usually leads to the same conclusion: that the favored hypothesis is established by the appropriate standard of proof.³³ Furthermore, excluding evidence where the likelihood ratio deviates substantially from 1:1 deprives the factfinder of information that might aid considerably in the rational resolution of disputed factual claims and may prevent a party from making what is, on a *fair* reading of all the evidence, a powerful case. This analysis supports the judicial practice of rarely, if ever, excluding evidence of substantial probative value simply because the jury appears likely to give the evidence even more weight than it deserves or because the precise weight to be given is unclear.³⁴ The preferred solution is to provide the jury with the information needed to assess accurately the probative value of the offered evidence.

A similar analysis applies where a court is called on to weigh the probative value of evidence against such factors as confusion of

31. Since there is good reason to believe that addicts often find it necessary to resort to crime in order to support their habits, if the assault were with an intent to rob the probative value of the evidence of addiction would, no doubt, be higher and the likely direction of the relationship would be clearer.

32. Other good reasons for this exclusion may also exist. *See, e.g.*, the discussion of prejudice in the text at notes 36-53 *infra*.

33. One can, of course, think of situations where the prior odds will be such that this argument does not hold. However, as a general matter trials are likely to be close enough that the analysis in the text applies.

34. There are other values that may justify the exclusion of highly probative evidence, *e.g.*, the rules of privilege and the rules regarding illegally seized evidence.

the issues, delay, and waste of time. Where the likelihood ratio for the evidence is far from 1:1, exclusion on these grounds is almost never justified except in the special case where, after considering all other admissible evidence, the court is convinced that the prior odds in favor of the disputed hypothesis are so high or so low that even highly probative evidence is unlikely to change the jury's judgment. This means that courts should be more reluctant in close cases than in clear ones to exclude probative evidence on such grounds as threatened delay, confusion, or waste of time. Appellate courts are certainly influenced by the closeness of cases in reviewing claims that the exclusion of evidence on such grounds was erroneous.

Each of those exclusionary rules that I call the relevance rules bars evidence of a particular type, but in most cases the bar is not complete. Exclusion is mandated only with respect to certain issues; on other issues the evidence remains admissible. Elsewhere I have argued that a general characteristic of the relevance rules is that the excluded evidence is rarely very probative of the issues on which it is inadmissible.³⁵ If this argument is correct, these rules of exclusion seldom force courts to contravene the policies advocated in the two preceding paragraphs. Instead they codify for recurring situations the decision rule that will usually be correct.

The Bayesian model that has been presented thus far aids in understanding the following aspects of the law relating to relevance: (1) the meaning of logical relevance, (2) the principle that only logically relevant evidence is admissible, (3) the discretion that courts have to exclude relevant evidence when the jury is likely to give it undue weight, (4) the reluctance of courts to exclude highly probative evidence although the jury is likely to give it undue weight, (5) the ways in which rules excluding certain evidence on specific issues relate to considerations of relevance, and (6) some of the justifications for those exclusionary rules that are generally seen as relating to relevance.

35. See R. LEMPERS & S. SALTZBURG, *supra* note 8, ch. 4. In some situations the excluded evidence will appear to have probative value. Where this is so it will almost always be the case that, if the hypothesis that the evidence is offered to support is true, there will be other available evidence that supports the desired inference even more strongly. In these circumstances, the other evidence will usually be sufficient to demonstrate by the appropriate standard of proof the truth of the hypothesis favored by the proponent of the evidence. The absence of such other evidence usually is an indication that on the facts of the particular case the inadmissible evidence has less than its usual tendency to prove the fact in dispute and suggests that the case presents one of those rare situations in which the inadmissible evidence is present although the hypothesis with which it is usually associated is not true. For an example of this, see the discussion of subsequent repairs in the text at notes 66-67 *infra*.

The Bayesian model does not, however, indicate why in some cases it might be desirable to exclude probative evidence not likely to raise estimation problems nor why it should be reversible error for a court to admit logically irrelevant evidence. However, another model drawn from decision theory helps clarify these aspects of the law of relevance. This model, called a *regret matrix*, aids in thinking about prejudice.

B. *Prejudice and the Regret Matrix*

A regret matrix³⁶ is not a normative model since it is not clear that the law expects the ideal decision maker to act in a manner consistent with it. It may, however, be a good descriptive model of the way decision makers, be they jurors or judges, actually behave, and values may be inserted into the model that are, arguably, normative. The model assumes that individuals wish to minimize the expected regret felt in the long run as a result of their decisions. In law, for example, a decision maker might wish to find for plaintiffs only when defendants were negligent. In terms of this model, the decision maker would have no regret in finding for plaintiffs when defendants were negligent and no regret in finding for defendants when they were not negligent.³⁷ Since in the uncertain world of litigation the decision maker can never be absolutely sure that a particular defendant was or was not negligent, the decision maker can never be absolutely sure of avoiding outcomes that would be regretted if the truth were known.

Although absolute certainty is impossible, the decision maker might be able to estimate a probability that the defendant was negligent, e.g., .6 or .7. If this can be done and if the decision maker can articulate the *relative* regret associated with different possible outcomes, a regret matrix can be constructed that indicates which decision—given the probabilities—leads to the least total regret in the long run. Consider the situation portrayed in Figure One.

36. What I shall refer to as a "regret" matrix is generally called a "utility" matrix in the decision theory literature. Since, as the matrix is used in this article, some disutility or regret is assigned to each of its various cells, I have followed the suggestion of Kaplan and call the matrix a "regret" matrix. See Kaplan, *supra* note 7, at 1078-82.

This use of the term "regret" should not be confused with the "risk" or "regret" payoffs associated with Professor L. Savage's minimax risk criteria of decision making that is applicable to decision problems in which the probabilities associated with various true states are unknown to the decision maker. See R. LUCE & H. RAIFFA, *GAMES AND DECISIONS* 280-82 (1957).

37. The example assumes that defenses such as contributory negligence are unavailable in this case, so liability turns solely on the issue of the defendant's negligence.

FIGURE 1

VERDICT	DECISION MAKER'S REGRET MATRIX		DECISION MAKER'S ESTIMATED PROB- ABILITY THAT D WAS:	DECISION MAKER'S EXPECTED REGRET IF VERDICT IS FOR:
	TRUE STATE OF AFFAIRS			
	D Negligent	D Not Negligent		
For P	0	1	Negligent .6	P .4
For D	1	0	Not Negligent .4	D .6

In this matrix no regret is associated with a decision for P when D was negligent or with a decision for D when D was not negligent. One unit of regret is associated with each mistake, that of finding for P when D was not negligent and that of finding for D when D was negligent. How should a decision maker with these values decide? That depends on his estimate of the probability that D was negligent. In the above example this probability is estimated at .6, making the estimated probability that D was not negligent (1 - .6) or .4. Knowing these probabilities, the expected regret for each verdict can be calculated by multiplying the regret associated with the verdict given the defendant's actual negligence or non-negligence times the probability that the defendant actually was negligent or not negligent. The sum of these products for a given verdict equals the total regret to be expected (in the long run) if that verdict were reached in all cases having the same regret matrix and probability of negligence. In the example, there is a .6 probability that D was negligent. Hence there is a .6 probability the decision maker who decides for P will feel no regret [$.6 \times 0 = 0$]. Conversely, there is a .4 probability that D was not negligent and that a decision for P will result in one unit of regret [$.4 \times 1 = .4$]. Thus, the regret expected from deciding for P given these probabilities of D's negligence will, in the long run, average .4 of whatever unit regret is measured in [$0 + .4 = .4$]. The situation is reversed when the decision is for D. There is a .6 probability that the decision maker will feel one unit of regret and a .4 probability that the decision maker will feel no regret. Consequently, the average expected regret from deciding for D is .6 units in the long run. An individual concerned with minimizing expected regret will decide for P in these circumstances.

The regret matrix used in this example is normative for most civil cases. A judge or juror *should* feel the same regret in reaching a mistaken decision for P that is felt in reaching a mistaken decision for D. If this is in fact the case (*i.e.*, if this particular regret matrix actually models the decision maker's values), one can show algebraically that regret is minimized by deciding for P whenever

the probability of negligence is greater than .5 and deciding for D whenever the probability of negligence is less than .5.³⁸

There are many civil cases in which a factfinder might feel uncomfortable with a norm that ascribes equal regret to the two kinds of mistakes. If this norm is rejected and if the factfinder seeks to minimize regret, he may strain to reach decisions that run counter to the weight of the evidence. For example, a juror whose insurance company connections make him sympathetic to tort defendants and hostile to injured plaintiffs might regret mistakenly deciding for P when D was not negligent twice as much as the opposite mistake. (This may be portrayed by changing the value in the upper right-hand cell of the matrix in Figure One to 2 while leaving the value in the lower left-hand cell at 1.) With this relative regret and the same probability that D is negligent as in the earlier example, .6 units of regret would be associated with a decision for D (the same as before) and .8 units of regret [$0 \times .6 + 2 \times .4$] with a decision for P. Hence a decision for D could be expected, although the decision maker's estimated probability of D's negligence is sixty per cent.³⁹

At law the burden of proof needed to sustain a conviction is the same for all defendants: good or evil, young or old, attractive or unattractive, dangerous or nonthreatening. Yet it is likely that jurors regret the mistake of convicting basically good people more than the mistake of convicting the basically evil. These feelings are reversed if the mistake is acquitting. The situation is undoubtedly similar with respect to other characteristics that affect people's attitudes toward their fellow human beings. If most jurors cannot avoid being influenced by such preferences in reaching their verdicts, the burden of proof is effectively changed by any information that affects these preferences. Consider, for example, the following regret matrices: one hypothesizes relative regret when a defendant is perceived as evil and the other relative regret when the defendant is perceived as good.

38. This is what is meant by a burden of proof by the preponderance of the evidence. Regret is equal when the probability of negligence is exactly .5. Here the law has decided that the defendant should prevail.

39. This assumes that a factfinder with the hypothesized regret schedule would be unwilling to accept the court's instruction that P should prevail if he establishes his case by a preponderance of the evidence.

FIGURE 2

		REGRET MATRIX FOR EVIL DEFENDANT		REGRET MATRIX FOR GOOD DEFENDANT	
		TRUE STATE OF AFFAIRS		TRUE STATE OF AFFAIRS	
VERDICT	Guilty	D Truly Guilty	D Truly Innocent	D Truly Guilty	D Truly Innocent
	Innocent	0	5	0	10
Innocent	1	0	1	0	

In the case of the evil defendant, the juror seeking to minimize regret would vote to convict whenever his estimated probability of the defendant's guilt exceeded .83. In the case of the good defendant the decision maker would require a probability of .91 before convicting.⁴⁰

The law's ideal juror estimates only the probabilities pertaining to the defendant's guilt and does not independently judge the regret associated with possible mistakes. This information is provided, in theory, by the court's instructions on the burden of proof. The requirement that guilt be proved beyond a reasonable doubt may mean that an accused should not be convicted unless the probability of guilt is judged to be at least .91, which is equivalent to saying that the law regards a wrongful conviction as being ten times more regrettable than a wrongful acquittal,⁴¹ or it may mean that conviction should

40. These probabilities are those which exist when the expected regret from the two possible verdicts (guilt and innocence) are equal. For the case where the regret associated with wrongful conviction is 5 times that associated with wrongful acquittal the appropriate equation is $0(X) + 5(1-X) = X + 0(1-X)$, which leads to $6X = 5$, or, after rounding, $X = .83$ where X is the probability the defendant is guilty. When the regret associated with wrongful conviction is 10 times that associated with wrongful acquittal the equation to be solved is $0(X) + 10(1-X) = X + 0(1-X)$, which leads to $11x = 10$, or, after rounding, $X = .91$.

41. This assumes that the two cells of the matrix which form what is called the "principal diagonal" are zero, *i.e.*, that no regret is associated with convicting a guilty person or acquitting someone who is innocent. It is possible that at least one of these two cells is not zero. For example, suppose that the decision maker believes that convicting a guilty person has regret or disutility associated with it because placing a person in the penal system is ultimately damaging to both the defendant and society. Preserving the perception that a wrongful conviction is ten times as regrettable as a wrongful acquittal, the following matrix might represent the relative regrets:

		TRUE STATE OF AFFAIRS	
		Guilty	Innocent
Verdict	Guilty	1	50
	Innocent	5	0

In this case, even though the relation between the regret associated with wrongful conviction and that associated with wrongful acquittal remains the same, the probability of guilt which one interested in minimizing expected regret would find necessary to convict is increased from .91 to .926. [$1(X) + 50(1-X) = 5(x) + 0(1-X)$, $50 = 54X$, $X = .926$]

It is likely that the regret matrices that would best model the behavior of actual

not follow unless some other minimum probability of guilt is obtained; but whatever the degree of certainty associated with proof beyond a reasonable doubt, the law does not contemplate that the standard of proof will vary with the defendant's personal characteristics or with the sordid details of the defendant's criminal activity.⁴²

In practice, the ideal of an unvarying standard is not achieved. Instructions on burden of proof, particularly in criminal cases, are so ambiguous that jurors necessarily exercise discretion in determining the degree of certainty needed to support a particular verdict. Furthermore, there is considerable evidence that jury verdicts are influenced by the personal characteristics of victims and defendants and by aspects of criminal activity that do not logically relate to the issue of guilt or innocence.⁴³ Where this occurs one may properly speak of prejudice, *for prejudicial evidence is any evidence that influences jury verdicts without relating logically to the issue of guilt or innocence.* Evidence that does relate logically to a disputed issue may also have a prejudicial effect, since the probative value of evidence may not fully determine its impact in the case. In terms of the regret model, one can conceptualize the prejudicial *potential* of evidence as the degree to which it affects the regret matrix of a juror viewing the case. The prejudicial *impact* of evidence depends upon prejudicial potential discounted by the juror's ability to ignore personal preferences in interpreting and applying the court's charge on burden of proof. Often the law fictively assumes that this ability is complete so long as the juror is instructed not to use evidence inappropriately. For simplicity's sake, this discussion shall

jurors would often contain nonzero values in the two cells of the principal diagonal. In this paper these values are set at zero because the ideal juror should feel no regret at reaching correct decisions. Indeed, evidence designed to make the factfinder regret correct decisions is typically considered irrelevant or prejudicial. A defendant, for example, could not introduce evidence of the conditions at the state prison to which he would be sent as part of his substantive case. Similarly it would be improper for a prosecutor to argue that the jury should hesitate to acquit an innocent person because an acquittal would give the impression that the jurisdiction was "soft on crime."

42. One might argue that the standard of proof should vary with certain characteristics of the defendant and that more doubt should be required to acquit the obviously evil or dangerous than to acquit the obviously good or nonthreatening. The argument, however persuasive it might be on the issue of how juries should behave given the interests of the larger society, is not relevant at this point in the analysis where I am treating law on its own terms, as an ideal system, in order to elucidate certain aspects of the law of relevance.

43. See, e.g., H. KALVEN & H. ZEISEL, *supra* note 14, at 301-47, 395-410. Relationships between verdicts and fact situations as reported in the Kalven and Zeisel study are consistent with the claim that jurors act, at least in part, to minimize personal regret.

proceed on the opposite and equally fictive assumption that prejudicial impact equals prejudicial potential.

The Bayesian model as it is used in this paper differs from the regret matrix used to model prejudice in an important respect. The discussion focusing on the Bayesian model assumes that the fact-finder assigns values rationally to the Bayesian equation. Either the values are assumed to be correct, as in the discussion of logical relevance, or errors are defined and attributed to a lack of information, as in the discussion of estimation problems. The regret matrix, as used here, admits of deviation from normative values but not of error and makes no assumptions about the rationality of the process that assigns values to the matrix. This process may be more or less logical, as when information of a defendant's felony record results in lowering the regret associated with the mistake of convicting because certain of the disabilities of conviction, most notably the stigma of a criminal record, already attach to the defendant. On the other hand, the process may be entirely devoid of logic, as where a change in regret results from evidence that is emotionally arousing, such as a gruesome photograph or an impassioned speech.⁴⁴

Much of the law relating to relevance reflects an awareness of the way in which prejudicial information can influence jury decision making. The danger of prejudice justifies the exclusion of some logically relevant evidence that does not pose estimation problems, and the same danger explains why the admission of logically irrelevant evidence may be reversible error. More specifically, a number of the relevance rules are justified in part because the evidence they exclude is fraught with prejudicial potential. I shall look at two of these rules by way of example.

44. It might be argued that with actual jurors the impact of emotion generally affects judgments that are analogous to those modeled by the Bayesian equation rather than by the regret matrix. For example, emotion may lead a juror to misestimate the probabilities in the likelihood ratio, or it may prevent a juror from "thinking straight" and thus lead to mistakes in calculation that make evidence appear more probative than it in fact is. If one could examine the psychological processes that underlie such mistakes, I would expect to find that they are most often made by jurors who wish to avoid confronting the fact that their relative regret is such that they are willing to convict on very flimsy evidence. In any case, the analytic value of the two-stage model proposed in this essay does not depend on the degree to which the two stages mimic the actual thought processes of jurors. Given the assumption that jurors seek to minimize personal regret, any mistake in a Bayesian calculation that would affect a juror's verdict can be portrayed as a change in the juror's regret matrix. This is analytically desirable regardless of actual decision processes, for it allows a clear separation between problems involving the probative value of evidence and problems that may arise because evidence can have an impact apart from its probative worth.

Consider first the rule that precludes introducing evidence of liability insurance to show negligence. The possession of liability insurance appears so unrelated to carefulness that a jury is not likely to treat the fact that a defendant was insured as tending to prove the defendant's negligence. Thinking solely in terms of Bayes' Theorem, evidence of the defendant's insurance coverage might be objectionable on the ground that its introduction wastes the court's time, but there is no reason to believe that such evidence will hurt either party. However, the regret matrix suggests a more substantial reason for excluding evidence of insurance. Knowledge that the defendant was insured may inappropriately affect the verdict whenever the factfinder's relative regret at mistakenly finding for or against an insured defendant will differ from the regret that would be felt if the factfinder thought the defendant would pay personally for the damages. Such a difference appears likely. Interestingly enough, some have argued that jurors should be informed of the existence of insurance because today's jurors assume insurance exists in all cases and construct their regret matrices accordingly. Insurance companies are not worse off when their interest in the case is revealed, so the argument goes, but uninsured defendants are harmed if jurors are not aware of their status.

The regret matrix also illustrates the sense behind rules like FRE 404(b) that forbid introducing evidence of other crimes, wrongs, or acts with the purpose of proving that a person acted in conformity with the character suggested by these delicts. While evidence that an accused committed some crime in the past may have probative value in that the probability of a history of crime may be higher for guilty defendants than for innocent ones, the probative value is likely to be outweighed by the effect that this knowledge will have on the jury's standard of proof.

If any regret matrix is normative for criminal cases it is probably the following:

FIGURE 3

		TRUE STATE OF AFFAIRS	
		D Truly Guilty	D Truly Innocent
VERDICT	Guilty	0	10+
	Innocent	1	0

This matrix is a mathematical portrayal of the oft-quoted statement that it is better that ten guilty men go free than that one innocent man be convicted. A substantial proportion of jurors may, in fact,

subscribe to approximately this norm.⁴⁵ Any evidence that leads a juror to change this initial regret matrix by diminishing his regret at convicting the innocent, raising his regret at acquitting the guilty, or associating regret with acquitting the innocent will prejudice the defendant in contravention of this arguably normative standard of proof. Evidence of other crimes is likely to do just this.

I have already alluded to one reason why such a change in the regret matrix might be expected from the revelation of a defendant's prior felonies: the stigma of being a felon attaches with the first conviction; subsequent convictions may also be stigmatizing, but they are not seen as having the same implications for a person's later life chances or his definition of self.⁴⁶ The other side of this is that the unblemished record, lost with the first conviction, has value in itself. Furthermore, some may regard a convicted felon as essentially criminal and believe that if he did not commit the crime charged he probably has committed or will commit other crimes. For these reasons the mistaken conviction of those with criminal records is likely to be perceived as less regrettable than the mistaken conviction of individuals thought never to have been in trouble with the law.

The danger of prejudice is likely to be even greater when it is shown that the defendant has engaged in illegal acts that have not resulted in criminal convictions. Here regret associated with a mistaken conviction is likely to be diminished because it is felt that the defendant deserves to be punished for his prior criminal activity whether or not he has committed the crime charged. For the same reason, regret, in contravention of the normative ideal, may be associated with the correct acquittal of such a defendant.⁴⁷ In addition to these difficulties, the attempt to prove guilt through evidence of bad character is fraught with estimation problems.⁴⁸

45. See Simon & Mahan, *Quantifying Burdens of Proof*, 5 LAW & SOC. REV. 319, 324 (1971).

46. See generally D. MATZA, *BECOMING DEVIANT* (1969); E. SCHUR, *LABELING DEVIANT BEHAVIOR: ITS SOCIOLOGICAL IMPLICATIONS* (1971).

47. The matrix that follows is an example of the kind of deviation from the normative that could be expected when a juror learns of the defendant's prior unpunished criminal activity.

		TRUE STATE OF AFFAIRS	
		Guilty	Innocent
VERDICT	Guilty	0	.5
	Innocent	1	.5

In the arguably normative matrix, figure 3 in text *supra*, the probability of guilt necessary to convict is .909. The effect of halving the regret associated with a mistaken guilty verdict is to decrease the requisite probability to .83. Associating regret in the magnitude shown with a correct verdict of innocent further reduces this probability to .818.

48. There is the "simple" problem of deciding exactly how character relates to

Of course these same problems exist when evidence of prior illegal or bad acts is admissible for permitted purposes, such as showing identity, opportunity, intent, motive, guilty knowledge, a criminal plan, or absence of mistake. Allowing the admissibility of other-crimes evidence to turn on the purpose for which the evidence is offered may be justified if such evidence has greater probative value and is less likely to raise estimation problems when it is offered for a permissible rather than an impermissible end. One can imagine situations where such evidence satisfies these criteria,⁴⁹ but whether this is the case generally is an empirical question that has not yet been answered. It does not, however, appear that even a colorable case can be made on relevance grounds for the common-law rule that allows evidence of *any* felony conviction to be admitted on the issue of a defendant's credibility.⁵⁰ The impeachment of non-party witnesses through evidence of other crimes is less objectionable because the evidence is not likely to affect substantially the relevant regret matrices. Moreover, if the evidence suggests that the witness is peculiarly susceptible to pressure from the state or the accused, as where a prosecution witness has been convicted but not yet sentenced, the evidence may be quite important in assessing the witness' credibility.

I have suggested that when a court weighs the probative value of evidence against such factors as delay, confusion, or waste of time, it should, in a close case, be reluctant to exclude the evidence if the

action and a more complex problem which exists because the defendant's record may have played a part in the decision to arrest and prosecute him. This latter problem will be discussed when I treat certain issues raised by the fact that evidence introduced at a trial is often not independent of other evidence introduced or of those factors that led the accused to be brought to trial in the first place. See text at notes 68-70 *infra*.

49. For example, in the famous "brides of the bath" case evidence of other drownings appeared highly probative on the issue of whether the drowning for which the defendant was tried was accidental. See NOTABLE BRITISH TRIALS, TRIAL OF GEORGE JOSEPH SMITH (E. Watson ed. 1922).

50. This rule might be explained as an historical anomaly or attributed to the tradition of not distinguishing for evidentiary purposes between defendants and other witnesses. A possible contemporary justification is that there may be a high probability that the truth will be distorted by anyone testifying in his own defense in a criminal case. This probability may justify putting a price on the decision to take the stand. But if the price—admission of evidence of previous criminal convictions—has no independent relevance on the issue of credibility, it does not seem fair that it be exacted only from those who have prior criminal records. If defendants without prior records are less likely to lie from the stand, but only because they are less likely to have committed the offense charged and so less likely to need to lie, the evidence is really relevant only for its bearing on character and the relationship of character to criminal activity, a purpose for which evidence of other crimes is generally not admitted.

likelihood ratio is much different from 1:1.⁵¹ One can easily imagine cases where evidence with a likelihood ratio of 2:1 or even lower would properly tip the balance.⁵² A court should be more willing to exclude probative evidence when it poses the danger of prejudice. It is easy to imagine evidence that might change a juror's regret matrix from the one that is arguably normative in criminal cases, *i.e.*, where mistakenly convicting generates ten times as much regret as mistakenly acquitting, to one in which the regret associated with these two mistakes is approximately the same.⁵³ In a close case, a change in juror regret of this magnitude could be devastating. Thus, a court is justified in excluding highly prejudicial evidence even if its probative value is substantial. However, it appears from the appellate cases that trial courts often refuse to exclude probative but prejudicial evidence. So long as the probative value of the admitted evidence is clear, appellate courts usually affirm such trial court decisions without attempting to weigh prejudicial effect against probative value.

II. RELAXING THE SIMPLIFYING ASSUMPTIONS IN THE APPLICATION OF BAYES' THEOREM

A. *Cumulative and Redundant Evidence*

Two simplifying assumptions are implicit in the portion of the preceding discussion that relies on Bayes' Theorem: (1) that the probative value of a given item of evidence may be determined with-

51. See text at note 34 *supra*.

52. If, for example, the prior odds in favor of the defendant's negligence were 3:5, a juror after receiving evidence supporting the defendant's negligence with a likelihood ratio of 2:1 would conclude that the odds were 6:5 that the defendant was negligent. Odds of 3:5 require a verdict for the defendant, while odds of 6:5 require a verdict for the plaintiff.

53. At this point, the analytic utility of the model does depend on the degree to which it represents, at least schematically, the actual behavior of jurors. It is possible that jurors can hear evidence that affects the regret they associate with possible verdicts yet neither distort the burden of proof as presented in the court's instructions nor make compensating distortions in their evaluation of the weight of the evidence. A juror's statement, "I would have preferred to convict the defendant, but given the evidence and the judge's charge I felt obligated to vote for acquittal" is not logically inconsistent, nor is it necessarily hypocritical. I believe that some jurors some of the time can separate their judgments about the desirability of convicting an accused from their obligation to render a fair verdict in accordance with the law and that many jurors most of the time can discount to some degree their own preferences in deciding a case. The discounting will not, however, be complete and in some cases may not occur at all. If these empirical hunches are correct, the situation depicted in the text may be rarer than one might intuitively guess, but it will occur, although frequently at levels of prejudice not as great as that suggested by the example.

out considering the other evidence in the case, and (2) that the reliability of evidence is not open to dispute. Relaxing these assumptions brings the Bayesian model into closer accord with the actualities of litigation and increases its utility as a heuristic device. The hypothetical trial that was the focus of the earlier Bayesian analysis consisted of only one item of evidence. Jurors receiving this evidence were expected to revise their odds on the defendant's guilt

in light of the likelihood ratio, $\frac{P(E|G)}{P(E|\text{not-G})}$. The likelihood ratio was used to explain the meaning of logical relevance and to explicate some of the concerns that courts have when dealing with problems of relevance.

Extending this model to trials involving two or more items of evidence would be straightforward if the revised odds of guilt after considering one item of evidence could be taken as the prior odds when considering the next item of evidence.⁵⁴ The extension is not this simple. Let us call the first item of evidence E_1 , the second item E_2 . Using $O(G|E_1)$, the odds on guilt arrived at after evaluating E_1 , as the odds on guilt existing before receipt of E_2 generally yields incorrect results except in the special case where E_2 is conditionally independent of E_1 with respect to the hypothesis of interest—*i.e.*, except where $P(E_1 \& E_2|G) = P(E_1|G) \cdot P(E_2|G)$ and $P(E_1 \& E_2|\text{not-G}) = P(E_1|\text{not-G}) \cdot P(E_2|\text{not-G})$.⁵⁵ These equations will not both be satisfied where part of the information conveyed by the presence of E_1 is taken into account when the implications of E_2 are evaluated.

54. The equations suggested by this straightforward extension, which in the general case is *not* correct, are

$$O(G|E_1) = \frac{P(E_1|G)}{P(E_1|\text{not-G})} \cdot O(G)$$

$$O(G|E_2) = \frac{P(E_2|G)}{P(E_2|\text{not-G})} \cdot O(G|E_1)$$

·

·

·

$$O(G|E_n) = \frac{P(E_n|G)}{P(E_n|\text{not-G})} \cdot O(G|E_{n-1})$$

where the first item of evidence is E_1 , the second item is E_2 , and the last item is E_n .

55. There are special cases where the chaining procedure illustrated in note 54 *supra* will yield correct results in the absence of conditional independence; *e.g.*, where $P(E_1|G) > 0$ and $P(E_2|G) = 0$.

An extreme example of redundant evidence should provide a clear illustration of why the straightforward extension of the Bayesian model is improper. Suppose in a murder case the factfinder at some point estimates the odds on the defendant's guilt as 1:100. The evidence that follows proves that the defendant's thumb print was found on the gun the killer used. The print of an innocent man might be found on a murder weapon because he handled the gun before or after the murder or because the print was planted there with the intention of framing him. Nevertheless, evidence of the print is surely more consistent with the hypothesis that the defendant is guilty than with its opposite. For the sake of this example assume that the factfinder believes that the presence of this evidence is 500 times more likely if the defendant is guilty than if he is not guilty.⁵⁶ Multiplying the prior odds of 1:100 by this likelihood ratio of 500 gives new odds on guilt of 5:1. Now suppose the prosecution wished to introduce evidence proving that a print matching the defendant's index finger was found on the murder weapon. If this were the only fingerprint evidence in the case, it would lead the factfinder to increase his estimated odds on the defendant's guilt to the same degree that proof of the thumb print did. Yet, it is intuitively obvious that another five hundredfold increase is not justified when evidence of the thumb print has already been admitted. This intuition is justified because having found the defendant's thumb print on the weapon, the probability of finding a print of the defendant's index finger if the defendant is guilty is not very different from the probability of finding this evidence if the defendant is not guilty. Thus, given the evidence of the thumb print, the likelihood ratio for the second fingerprint is approximately one.⁵⁷

56. A mathematically inclined juror might, for example, believe that there is a .2 probability that the print would be found if the defendant were guilty (the probability is considerably less than one because guilty people often have taken the trouble to wipe their prints from weapons and, even if they had not, not all prints are identifiable) and a .0004 probability that the evidence would be found if the defendant were not guilty. Note that later evidence, suggesting a plausible reason why the defendant, although innocent, might have left his prints on the gun, could substantially increase this probability, thus leading the juror to reduce substantially the probative weight accorded this evidence.

57. The presence of the second print depends largely on the way the defendant held the gun when he left the thumb print. Unless murderers hold guns differently than nonmurderers or are more likely to wipe off some but not all their fingerprints, the finding of the second print is no more consistent with the hypothesis that the defendant is guilty than with its opposite. Indeed, because a murderer is more likely to attempt to wipe off fingerprints from a gun than one with no apprehension of being linked to a murder and since an attempt to wipe off fingerprints might be only partially successful, there is a plausible argument that the presence of the second print should lead jurors to be somewhat less confident that the defendant is the murderer than they would be if only one of the defendant's fingerprints were found.

Where items of evidence are not independent, the simplest way to apply the variant of Bayes' Theorem that we have been using is to treat the interdependent evidence as a single event when calculating the likelihood ratio. Thus for any two items of interdependent evidence the likelihood ratio equals:

$$\frac{P(E_1 \& E_2|G)^{58}}{P(E_1 \& E_2|not-G)}$$

This procedure, conceptualizing two items of evidence as a single item, may be generalized to account for interdependence among any number of items of evidence, generating a likelihood ratio of the form:

$$\frac{P(E_1 \& E_2 \& \dots \& E_n|G)}{P(E_1 \& E_2 \& \dots \& E_n|not-G)}$$

Since the analysis for more than two items of evidence is basically the same as for two items, this article will focus on the two-item case to simplify the discussion.

Extending the model in this way makes it a useful device for exploring ways in which evidence may be cumulative. An item of evidence, E_2 , introduced after some other item, E_1 , is properly considered cumulative and may be excludable for that reason when the

ratio $\frac{P(E_1|G)}{P(E_1|not-G)}$ is identical with or very close to the ratio

$$\frac{P(E_1 \& E_2|G)}{P(E_1 \& E_2|not-G)}$$

Where this condition is met, consideration of the second item of evidence, E_2 , adds little or nothing to what may be learned from the proper consideration of the first item, E_1 .

Another variant of Bayes' Theorem, expressing the chance of guilt in terms of probabilities rather than odds, is helpful in specifying when the likelihood ratio for E_1 will be close to or identical with the likelihood ratio for E_1 and E_2 taken together. For the case where there are two pieces of evidence, Bayes' Theorem may be expressed as:

58. This likelihood ratio is correct for any two items of evidence, whether or not they are interdependent. However, where the items are conditionally independent, this ratio will be equal to the product

$$\frac{P(E_1|G)}{P(E_1|not-G)} \cdot \frac{P(E_2|G)}{P(E_2|not-G)}$$

This result follows immediately from the equations in the text at note 55 *supra*. When the items are conditionally independent, the extension of the Bayesian model suggested in note 54 *supra* is appropriate.

$$P(G|E_1 \& E_2) = \frac{P(E_2|G \& E_1)}{P(E_2|E_1)} \cdot P(G|E_1)^{59} \quad (4)$$

When the ratio $\frac{P(E_2|G \& E_1)}{P(E_2|E_1)}$ equals or is close to one the likelihood ratios for E_1 and for E_1 and E_2 taken together are identical or virtually so.⁶⁰ This situation exists when E_2 adds little if any information concerning the probability of the defendant's guilt to that which was provided by E_1 . In these circumstances the likelihood of finding the second item of evidence, E_2 , depends upon the relationship between E_2 and E_1 rather than on whether the defendant is guilty. In the example of the fingerprints, once it is known that the defendant's thumb print was left on the gun the estimated probability of finding another of the defendant's fingerprints there will not change with knowledge of the defendant's guilt. The reason for this is that finding the thumb print related to the defendant's guilt only insofar as it proved that the defendant handled the gun, and finding another fingerprint relates to the defendant's guilt in exactly the same way. Since the second print proves a fact already established it is redundant.

A second situation in which $\frac{P(E_2|G \& E_1)}{P(E_2|E_1)}$ equals one is when E_1 is sufficient to establish G , since where this is true, knowing E_1

59. This variant of the Theorem is derived through successive applications of equations (1) and (3) of note 12 *supra*:

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

60. Where $\frac{P(E_2|G \& E_1)}{P(E_2|E_1)} = 1$, the formula in the text at note 59 *supra* indicates that

$P(G|E_1 \& E_2) = P(G|E_1)$. Application of equation (3) in note 12 *supra* results in

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

which can be shown to imply

$$\frac{P(E_1 \& E_2|G)}{P(E_1 \& E_2|not-G)} = \frac{P(E_1|G)}{P(E_1|not-G)}$$

i.e., the likelihood ratios are identical when

$$\frac{P(E_2|G \& E_1)}{P(E_2|E_1)} = 1.$$

means that one also knows G .⁶¹ Thus, the numerator and denominator of the ratio are identical.

The Bayesian analysis demonstrates that where the likelihood ratios for E_1 and for E_1 and E_2 taken together are identical or virtually so, consideration of the second item of evidence adds little or nothing to what may be learned from a proper evaluation of the first item. Nevertheless, there are situations in which a court should admit evidence that is analytically cumulative in this sense. The first such situation is where it is possible that the jury does not appreciate fully the information conveyed by the first item of evidence. For example, suppose that a gynecologist was accused of participating in a criminal abortion. Evidence that cervical dilators, instruments used with any of the standard techniques of early abortion, were found in the gynecologist's office would be cumulative once the defendant's profession was shown, for the ordinary practice of gynecology requires a physician to have these instruments available. However, the jury might not realize that all gynecologists have access to cervical dilators, so the prosecution should be allowed to introduce evidence showing access in the particular case. In Bayesian terms, $P(E_2|G \& E_1)$, the probability of finding E_2 (evidence of the dilators), would be perceived as one if the defendant were guilty and a gynecologist (E_1), but $P(E_2|E_1)$, the probability of finding the dilators knowing only that the defendant was a gynecologist, would be misperceived as less than one. Thus, a juror would attach different

values to the ratios $\frac{P(E_1|G)}{P(E_1|\text{not-G})}$ and $\frac{P(E_1 \& E_2|G)}{P(E_1 \& E_2|\text{not-G})}$ so the evidence of the dilators would not be perceived as cumulative in the context of the case.

There is, however, the danger that admitting evidence of the cervical dilators would raise another estimation problem. Jurors who know little about gynecology might not find evidence of the defendant's profession very probative of guilt, but they might find the defendant's possession of the cervical dilators highly so. In this situation, the defense counsel must attempt to put the evidence in context by showing the uses of these instruments in the ordinary practice of gynecology. If both the prosecution and the defense counsel do

61. As a practical matter one may point to a third situation where courts may be justified in rejecting evidence as cumulative even though this ratio does not equal one. This is where E_1 , although not sufficient to establish G to a certainty, is sufficient to increase the probability of G far beyond what is required for conviction beyond a reasonable doubt. The discussion of the second situation described in the text applies generally to this third situation, also.

their jobs properly, the jurors should be able to estimate the true probative value of the evidence that the defendant is a gynecologist, and they should realize that once this fact is taken into account the defendant's possession of cervical dilators adds nothing.

A second situation in which cumulative evidence should be admitted is where the jury expects that the evidence will be produced if it exists. The absence of evidence conveys information to the jury, and it is possible for the proven availability of evidence to be cumulative while its proven unavailability has considerable probative value. In these circumstances cumulative evidence should be admissible, despite slight probative value, in order to dispel the implication that it is unavailable. An example should make this clear. Consider a murder trial in which the following facts have been established and weighed by the jurors in setting their odds on guilt: (1) the victim was killed by a shotgun, and (2) the killer is a resident of a particular community, 99% of whose residents have access to shotguns. Once E_1 , that the defendant is a resident of the suspect community, has been established, the further evidence E_2 , that the defendant has access to a shotgun, does little to increase the estimated probability of the defendant's guilt. Since it is certain that the defendant will have access to a shotgun if he is both the killer and a resident of the suspect community, and there is a .99 probability that he will have access to a shotgun if he is a resident of the community, the Bayesian equation is⁶²

$$P(G|E_1 \& E_2) = \frac{1.0}{.99} \cdot P(G|E_1).$$

Thus, specific evidence that a shotgun was available to the defendant raises the previously estimated probability of the defendant's guilt by only about one per cent. Yet failure to introduce evidence of ownership might harm the prosecution's case, since the jury might treat the absence of such evidence as a fact having probative value. Indeed, for a juror who was certain that the prosecution would introduce evidence of the defendant's access to a shotgun if the defendant had access, the lack of evidence would be reason to acquit, for E_2 would be "evidence of no access" and the appropriate equation would be

$$P(G|E_1 \& E_2) = \frac{0}{.01} \cdot P(G|E_1).$$

62. Equation (4), the variant of Bayes' Theorem developed in the text at note 59 *supra*, is used here.

Thus, whatever the prior odds of guilt, the failure to produce evidence of access is exonerative. A juror with this perspective would be confused, for the prosecution's failure to show that the defendant is a shotgun owner is not equivalent to proof by the defendant that he is not. Nevertheless, some jurors might be confused in this way. Furthermore, failure to prove ownership is more consistent with the hypothesis that the defendant is not guilty than with its opposite, so even a juror who was not confused would be justified in lowering his odds on the defendant's guilt if the defendant's ownership of a shotgun were not shown. The lesson of this example is that there are times when a party may properly insist on the admission of cumulative evidence in order to dispel unwarranted inferences about its unavailability.

Ordinarily the admission of cumulative evidence will not be reversible error, since cumulative evidence by definition does not affect the rational factfinder's judgment of the odds on the defendant's guilt. Indeed, if the costs of presenting cumulative evidence are not great, the decision to allow cumulative evidence is often wise. Modes of proving the same fact differ in the degree to which they command attention and in the likelihood that they will be understood. Thus a juror might appreciate the implications of certain evidence without realizing that the information conveyed by that evidence was implicit in earlier proof. Of course, when a fact is being proved for a fifth or sixth time the benefits of redundancy are likely to be low.

The admission of cumulative evidence may be reversible error where the evidence is prejudicial. Since cumulative evidence ordinarily is not needed to establish a point, its probative value will be outweighed by even a slight possibility of prejudice.⁶³ However, the problem a trial judge faces is not as simple as this analysis might suggest, for evidence that appears cumulative when offered may not appear cumulative after the opposing side has cross-examined a witness or presented its case. One solution is to require the opposing party to stipulate to the proposition that the evidence tends to establish before excluding it as cumulative. Another solution is temporarily to exclude the evidence, subject to an opportunity to reoffer it if later evidence suggests that it is not, in fact, cumulative. Where

63. If the evidence is cumulative (in the sense that it is further proof of something that has been indisputably established) on the ultimate issue in the case (*e.g.*, guilt) rather than on some constituent fact, a mistaken decision to admit prejudicial evidence should be harmless error, since even without the evidence a reasonable jury would not have reached a different conclusion.

the potential for prejudice is substantial rather than slight, outright exclusion is often justified.

Special problems exist where several items of evidence persuasively prove the same fact and nothing more. In this situation courts are well advised to allow proof only by the least prejudicial evidence even though the evidence first offered to prove a fact cannot be considered cumulative. Many courts, however, are reluctant to interfere with a party's chosen mode of proof. Thus, in homicide cases prosecutors have been allowed to prove the fact of the victim's death by gory photographs despite the availability of less emotive testimony or, even, a stipulation.⁶⁴ This reluctance to interfere with a party's preferred form of proof exists even when the preferred proof follows the admission of less prejudicial evidence and so is clearly cumulative. Evidence of other crimes, for example, has been received for its bearing on issues not actually in dispute or on facts provable by the overwhelming weight of other evidence.⁶⁵

Several of the relevance rules may be justified, in part, because the evidence they exclude is likely to be either cumulative or of low probative value. Evidence of subsequent repairs, for example, is by rule inadmissible on the issue of negligence.⁶⁶ Even if this evidence generally has some tendency to prove negligence, the rule of exclusion is unlikely to harm deserving plaintiffs. Where the defendant was, in fact, negligent, other more probative evidence is likely to be available. Thus, in a typical well-founded negligence action, evidence of subsequent repairs is likely to be cumulative on the central issue. On the other hand, where the only evidence of negligence is a subsequent repair, it is unlikely that the plaintiff had a valid claim to begin with since the probability of finding this evidence and no other is probably less in the case of negligent defendants than in the case of nonnegligent defendants.⁶⁷

64. Courts differ in this respect. Some cases do hold that it is an abuse of discretion to admit potentially prejudicial evidence when a party has offered to stipulate to everything that evidence might legitimately be admitted to prove. *See, e.g., Note, Inflammatory Photographs: How Sensitive Are Texas Courts to Unfair Prejudice?*, 29 BAYLOR L. REV. 154 (1977).

65. *See, e.g., State v. Goldberg*, 261 N.C. 181, 134 S.E.2d 334, *cert. denied*, 377 U.S. 978 (1964).

66. FED. R. EVID. 407.

67. Even if this is the case, one might ask why evidence of subsequent repairs should be excluded, since this type of evidence is not likely to prove prejudicial. My answer is that this rule, like most of the relevance rules, has multiple justifications. I believe the likelihood of low relevance is crucial because it frees courts to look at other considerations. In the case of the rule regarding subsequent repairs, the classic other justification is that the decision to admit such evidence would be

When evidence is excludable as cumulative, the factfinder has considered all of its relevant informational content in his weighing of other evidence. It is also possible for evidence to be in some degree redundant without being so redundant as to be cumulative. Evidence of this sort is admissible subject to the court's ordinary discretion to weigh probative value against such factors as prejudice, confusion, and waste of time. The major danger in admitting partially redundant evidence is the possibility of an estimation problem: the jurors may not appreciate the redundancy; thus they might give the evidence more weight than it deserves.

Recall that the Bayesian model presumes that the factfinder approaches new items of evidence with some estimate of the prior odds that the defendant is guilty. These odds change as relevant evidence is received, and each subsequent item of evidence is evaluated with respect to the most recent estimate of the prior odds. When evidence is cumulative or redundant, part or all of the informational content of the evidence has been considered in setting the odds that exist prior to the receipt of that evidence. In this situation, it is the task of counsel to put the evidence in its proper context, one that suggests the evidence is not as persuasive as it might otherwise appear. Otherwise there is the likelihood that certain aspects of the evidence will be counted twice.

When separate items of evidence are introduced the possibility of redundancy is generally clear. What may not be obvious is that information conveyed by an item of evidence may be redundant even though previous evidence conveying that information has not been received. If the Bayesian model approximates the process by which people evaluate evidence, a factfinder must begin with the belief that there is some chance that the defendant is guilty because once the estimated prior odds on guilt are zero no amount of subsequent evidence, however persuasive, will change the rational factfinder's evaluation of those odds. One might argue that the presumption of innocence means that at the commencement of a case the factfinder's estimated odds on guilt should be one to whatever figure represents

harmful because it would discourage the making of repairs after accidents. Elsewhere I suggest that this justification is vulnerable to criticism because of the assumption it makes about knowledge of the rule. There I suggest another possible justification: recognition that a defendant who has repaired a hazardous condition has taken a socially responsible action and should not be made to suffer, or even to appear to suffer, for this. Evidence of subsequent repairs is admissible on issues other than negligence because as its relevance becomes greater such delicate considerations as this one must give way to the need to get at the truth. See R. LEMPERS & S. SALTZBURG, *supra* note 8, at 186-89.

the size of the relevant population from which the defendant came.⁶⁸ However, the little empirical evidence that exists suggests that jurors begin with much higher estimated odds on guilt, sometimes as high as 1:1.⁶⁹ In part, this is because jurors quite rationally assume that the defendant would not be before them if there were not special reasons to suspect him of the crime. Where this assumption has played a part in setting the initial odds, some part of the evidence that was crucial to the police's decision to arrest and the prosecution's decision to proceed to trial has been implicitly counted against the defendant before its presentation to the jury. Since it is not clear how much double counting has occurred or which evidence is most likely to be redundant, controlling for this possibility is difficult, if not impossible.⁷⁰

One class of evidence poses special dangers of double counting. This category consists of evidence that relates more strongly to the probability that the defendant would be arrested than to the probability that he committed the crime in question. Most likely to fall into this category is evidence relating to the defendant's prior record. When the perpetrator of a crime is unknown, police commonly focus their attention on individuals known to have committed similar crimes in the past. Furthermore, when a photographic identifica-

68. John Kaplan has suggested that from a normative perspective the prior odds on a defendant's guilt should be one to about 200,000,000, since that is the approximate population of this country. Kaplan, *supra* note 7, at 1085-86. If a juror began with these prior odds, he would be justified in revising them drastically downward upon receipt of evidence proving such facts as the defendant's age, if it were such as to indicate that the defendant was capable of committing the crime; the defendant's place of residence, if it placed him in a locale convenient to the commission of the crime; the defendant's race, if the criminal had been identified by race and the defendant was of the criminal's race; etc. More realistic prior odds would probably take these kinds of factors as implicit in the defendant's arrest and thus be set at one to whatever figure represents the number of people possessing those gross characteristics that were virtually certain to characterize anyone arrested for the crime.

69. See, e.g., Weld & Roff, *supra* note 16, at 617.

70. It is unclear how great a danger exists, since the extent of the danger depends upon the way jurors in fact behave. Even though jurors will report a prior probability of guilt to researchers after hearing an indictment read, *see id.* at 617, this may be a very tentatively held prior probability, subject to revision after some evidence has been presented. Also, jurors may treat a failure to produce evidence as itself evidence, *see* text following note 62 *supra*, particularly where an assumption about the availability of that evidence entered into the initial estimation of the prior odds. Finally, certain evidence that entered into the estimation of the initial prior probability may be received by the jury during the course of the trial without resulting in double counting. For example, a juror may say that he did not increase his estimate of the defendant's guilt upon learning that the defendant's race, sex, or hair color matched those of the criminal because he assumed that the police would not have arrested anyone unless these characteristics were consistent with the known characteristics of the criminal.

tion is sought, photos may be available only for persons with criminal records. These considerations suggest further support for the rule excluding evidence of past crimes where the evidence is relevant only insofar as it suggests the defendant has a propensity toward crime. Such evidence is likely to be redundant because it probably has influenced the decisions to arrest and prosecute and so has figured in the jurors' initial estimation of the odds on the defendant's guilt. Even if jurors could be made aware of the way in which this type of evidence enters into their initial estimation of the odds on guilt, the task of separating the redundant portion of such evidence from the nonredundant poses an insoluble estimation problem. Thus, there is sense in the prevailing view that evidence of other crimes should be dealt with by a rule of exclusion rather than on an ad hoc basis.

B. *Reliability of Evidence*

In determining whether evidence is redundant to the point of being cumulative, attention must be paid to the source of the evidence as well as to its substantive content. Thus far the discussion has taken the reliability of evidence as given and has assumed that the rational factfinder's only task is to compare the probability that certain facts could be proved if the defendant were guilty with the probability that the same facts could be proved if the defendant were innocent. The oversimplification is obvious. The jury does not typically hear proof of facts; rather, it hears testimony tending to prove certain facts or receives evidence that arguably bears on the case but is not indisputably linked to it. Upon hearing testimony, jurors must compare the probability that the testimony *would be given* if the defendant were guilty with the probability that the testimony *would be given* if the defendant were innocent. This comparison involves an estimate of the reliability of the testimony as well as an estimate of the probative worth of the facts that the evidence tends to prove.

Because the jury is faced with this twofold task, impeachment evidence serves an important function. The likelihood ratio based on the facts that the testimony tends to prove sets the upper bound on the probative value of that testimony. Any evidence suggesting that the testimony may be inaccurate will lead a factfinder to decrease his estimated likelihood ratio.⁷¹ Cross-examination holds its

71. Rarely will impeachment evidence be such that the likelihood ratio for the impeached testimony dips below 1:1 on the hypothesis that the facts that the testimony tends to prove are true. Thus, the general rule that a party may not rely solely on the jury's disbelief of an opposing witness to prove an element in his case is a sound one. It should also be noted that an argument analogous to that in the text applies where the authenticity of real evidence is in dispute.

exalted place in the Anglo-American system of trial procedure because it is thought to be an effective means for exploring the possible inaccuracies of testimony and, hence, a valuable aid to rational fact-finding.

Since human error is always possible when witnesses report events, it is rarely if ever proper for a court to exclude the testimony of one witness as cumulative simply because another witness has testified to the same fact. At some point, however, similar testimony by additional witnesses will be cumulative. That point will depend upon the degree to which the testimony of earlier witnesses has been challenged and the way in which different testimony is vulnerable. If the testimony of several witnesses has not been challenged, little reason exists for allowing other witnesses to testify to the same point.⁷² Where the testimony of earlier witnesses has been challenged, the testimony of later witnesses is unlikely to be cumulative unless open to identical challenge.

As an illustration, consider the likelihood ratio:

$$\frac{P(T_1 \& T_2 \& \dots \& T_n | G)}{P(T_1 \& T_2 \& \dots \& T_n | \text{not-G})}$$

where T_1, T_2, \dots, T_n is testimony from a series of witnesses, W_1, W_2, \dots, W_n . Let us suppose this testimony identifies the defendant as a car thief. If W_1 knew the defendant well but was embroiled in a family feud with him, the factfinder might, after hearing vigorous cross-examination, believe that the likelihood ratio for T_1 taken alone was close to one. W_1 would surely accuse the defendant of theft if he knew the defendant was guilty, but, because of his family's feud with the defendant, he might be almost as likely

72. A potential problem exists because a party might, without overtly challenging opposing testimony, introduce his own witnesses to contradict that testimony. As a practical matter, however, it is unlikely that a party who wished to dispute the story of opposing witnesses would entirely forgo cross-examination.

The desire to avoid cumulative testimony does suggest one justification for the rule forbidding certain types of impeachment by extrinsic evidence unless a foundation has been laid on cross-examination. The knowledge that a party intends to impeach particular witnesses aids a court in determining whether further testimony along certain lines is likely to be cumulative. I do not suggest that this justification, either alone or in combination with other justifications, is sufficient to override the case that can be made for dispensing with a foundation in some of the situations where the common law now requires one. For example, the rule requiring that a witness be confronted with his earlier inconsistent statements before the statements may be proved extrinsically certainly undercuts some of the dramatic force of impeachment. Whether the jury gets a better picture of the witness' trustworthiness from seeing the witness react to an inconsistent statement before it is proved extrinsically or whether they form a more accurate judgment by hearing the statement subject to the witness' later opportunity to explain or deny it is an empirical question. I feel less confident that I know the answer to this question than many who have written on this subject.

to accuse the defendant if the defendant was innocent. If W_2 , W_3 , and W_4 were members of W_1 's family who claimed to have witnessed the same theft, their testimony would be vulnerable in the same way as W_1 's testimony and thus add little to the prosecution's case. This additional testimony would not be completely irrelevant because the fear of prosecution for perjury and the innate honesty of the four witnesses might differ, even though they all shared the same interest in seeing the defendant convicted. However, at some point the court would be justified in heeding the defendant's argument that there was little more to be learned from further identification by other members of W_1 's family. The jury is likely to have reached a judgment about the probability that everyone in W_1 's family would accuse the defendant of theft if he was guilty and about the probability that there would be a similar unanimity of accusation if the defendant was innocent. Further accusations from other family members are unlikely to change this judgment.

On the other hand, the testimony of a second witness who had no quarrel with the defendant but was nearsighted, of a third who caught only a fleeting glimpse of the defendant, and of a fourth who had seen the defendant for about a minute but had a poor memory for faces would not be cumulative, or even redundant, except in its tendency to prove the identity of the car thief. Even if the testimony of each of these witnesses, taken alone, has little probative value because of its peculiar weaknesses, the testimony of all four witnesses has substantial probative value since the probability of the same mistake being made by four witnesses with such different reasons to err appears small.⁷³ To generalize, where the testimony of two or more witnesses tends to prove the same fact, the relevance of the later testimony will be greater (1) the less the apparent reliability of the earlier testimony and (2) the greater the variability in the reasons for doubting the stories of the different witnesses.⁷⁴

73. It is also the case that in balancing relevance against the potential for prejudice, confusion, or waste of time, a court should weigh the evidence in the context of other evidence in the case rather than as a discrete item. Sometimes the evidence will be less relevant when taken in context while on other occasions it will be more so.

74. This argument applies as well when only real evidence is involved or when a combination of real evidence and testimony is presented. For example, where a series of crimes has occurred that is so distinctive that the same individual almost certainly committed all of them, the relevance of the other-crimes evidence will depend upon the strength of the evidence linking the defendant to each crime and the independence of this evidence for each crime. If, for example, it was proved that all the crimes were committed with the same gun and the defendant was arrested

This analysis suggests strategies for seeking and presenting evidence. In attempting to prove a disputable point, an attorney should seek items of evidence that do not share the same sources of possible unreliability. In attempting to destroy an opponent's case, counsel should strive to show that the evidence of the opponent is infected from a common source. In the hypothetical case of the car thief presented above, the prosecutor should, after finding several members of W_1 's family willing to testify against the defendant, expend his limited resources searching out witnesses who are not members of W_1 's family rather than looking for more witnesses who are parties to this feud. A defense counsel faced with a varied array of identification witnesses should try to show that their identifications share a possible source of error, as would be the case if before identifying the defendant each witness had been shown the defendant's picture and told that the police thought he was the thief. In an attempt to establish a common source of error, counsel offers an explanation for the testimonial coincidence other than the fact that the witnesses are all responding to the same initial stimulus.

III. CONCLUSION

I assert in the introduction that mathematics, as a language, can help clarify those legal rules that involve the weighing of evidence in an essentially probabilistic fashion. This article proceeds on the assumption that the rules relating to relevance are such rules. I believe this to be the case, but I also believe that there are aspects of, and justifications for, the relevance rules that have nothing to do with the rational evaluation of evidence. As with almost any other area of law, values that defy quantification must be attended to in

with that gun in his possession but charged only with the last of the series of crimes, evidence of the earlier crimes would be irrelevant on the issue of whether the defendant committed the crime charged. Whatever excuse the defendant could give concerning his possession of the gun (*e.g.*, he had recently bought it from a friend) would, if accepted, destroy the link between the defendant and each of the other crimes. The evidence of other crimes would, however, be prejudicial, since the jury might regret the mistake of acquitting one who, if guilty of any crime, is guilty of several, more than they would regret the mistake of acquitting one thought to have committed at most one crime. If, on the other hand, the defendant was linked to the other crimes by the identification testimony of different victims, the evidence of the other crimes would be relevant. Even if the defendant could show weaknesses in the testimony of one of these witnesses, it is unlikely that he could show that the testimony of all the witnesses was incredible or was vulnerable to the same objection. Prejudice would still exist here, but there would be substantial probative value since, by hypothesis, it is reasonably certain that the same person is responsible for all of the crimes. Thus the prejudicial impact of this evidence might be outweighed by its probative value, and a court might be justified in admitting it.

analyzing relevance. Any mathematical treatment is necessarily limited.

The mathematical models used in this article do not by themselves answer fundamental policy questions. Their purpose is to stimulate insight and to aid in the clear and concise explication of what is perceived. Although I do not think that jurors in actual cases should be urged to use Bayesian calculations in evaluating evidence, nevertheless, I believe the approach taken here is of more than strictly academic interest. Whatever enables lawyers to think more clearly is of practical importance. When faced with difficult problems of relevance or harmless error, attorneys may find that the models presented aid in thinking about the implications of admitting certain evidence. Using the models requires specification of the probabilities that make up the likelihood ratio for the evidence and a determination of possible prejudicial impact. It may also require attorneys to estimate the jury's likely evaluation of the evidence and the probative value of the other evidence in the case. Typically it will be impossible to obtain general agreement on the specific values to be incorporated into the mathematical analysis, but it may be possible to gain agreement on upper or lower bounds for these measures. If either model suggests that an argument is valid when the values incorporated into the model are at the extreme least favorable to the claim advanced, the argument is almost certain to be valid for the evidence in the case. Where this is so, an attorney has a precise and powerful means of arguing to any judge who can be persuaded to think in terms of these models.

Much of the analysis in this article does not depend upon the extent to which the models used portray the ways in which jurors actually respond to evidence. The Bayesian model is normative—it specifies the way in which jurors are expected to evaluate evidence—and values may be inserted into the regret matrix that are arguably normative. Since these models describe the behavior of the ideal legal factfinder, they allow us to evaluate aspects of legal factfinding in terms of the ideal as well as to speculate on how actual factfinding deviates from the ideal. At some points, however, the argument does depend on the degree to which these models portray actual juror behavior. This was typically the case where I went beyond explanation and drew on these models to suggest how courts *should* respond to certain kinds of evidence. We do not know the extent to which jurors process evidence in a Bayesian fashion, nor do we know precisely how jurors are influenced by the regret that they are likely to associate with alternative verdicts. These are important areas for

further research.⁷⁵ If empirical research should reveal that jurors act, more or less, as the two models suggest, the utility of approaches like the one taken in this article will be substantially enhanced.

75. For a recent review of some of the literature on people as decision makers, see Slovic, Fischhoff, & Lichtenstein, *Behavioral Decision Theory*, 28 ANN. REV. PSYCH. 1 (1977).