



Burdens of Persuasion: What Bayesian Decision Rules Do and Do Not Do

D.H. Kaye⁽¹⁾
Arizona State University
Tempe, AZ 85287-7906

A final version of this paper is published in the *International Journal of Evidence and Proof*, Vol. 3, No. 1, 1999, pp. 1-28

--Probability, like logic, is not just for mathematicians anymore.⁽²⁾

--Condemned to the use of words, we can never expect mathematical certainty from our language.⁽³⁾

Everybody knows that the prosecution in a criminal case has the burden of proving its case beyond a reasonable doubt. Every lawyer knows that the plaintiff in a typical civil case has the burden of proving its case by a preponderance of the evidence. But what do these hoary phrases mean? That the probability p of a set of facts giving rise to civil liability must be greater than 50%? That p for criminal liability must exceed 95%? Why this difference (or any other)? To answer questions like these, legal scholars have applied the tools of statistical decision theory to legal factfinding. They have reached the following, sometimes surprising results:

- *The expected-utility-maximizing rule for a simple civil case:* In a case in which a plaintiff who suffered an injury from either one defendant or some other cause sues that defendant, the $p > \frac{1}{2}$ rule maximizes expected utility when the utility of factually correct verdict for plaintiff equals the utility of a factually correct verdict for defendant, and the disutility of an incorrect verdict for plaintiff equals the disutility of an incorrect verdict for defendant.⁽⁴⁾
- *The expected-utility-maximizing rule for a simple criminal case:* In a case in which the utility of a factually correct guilty verdict equals the utility of a factually correct acquittal and the disutility of a false conviction exceeds the disutility of a false acquittal, a conviction maximizes expected utility when p exceeds some level p^* that is itself greater than $\frac{1}{2}$ and that depends on the ratio of the two disutilities.⁽⁵⁾
- *The expected-error-minimizing rule for a simple civil case:* In a case in which a plaintiff who suffered an injury from either one defendant or some other cause sues that defendant, the $p > \frac{1}{2}$ rule minimizes the expected number of factually incorrect verdicts.⁽⁶⁾
- *The expected-error-minimizing rule for a single-cause, multiple-defendant civil case.* In cases in which a plaintiff suffered an injury from exactly one of several defendants, the rule that assigns liability to the single defendant who most probably caused the injury (even if $p > \frac{1}{2}$ for this defendant) minimizes the expected number of dollars paid by parties who are not responsible for the loss.⁽⁷⁾
- *The expected-error-minimizing rule for a single-cause, multiple-plaintiff civil case.* In cases in which many plaintiffs suffered harm from either the defendant or other possible causes, and the probability of defendant's responsibility varies among the plaintiffs, but the defendant caused a known amount of damages to some proper subset of the plaintiffs, the "most likely victim" rule minimizes the expected number of dollars paid to plaintiffs who were not injured by defendant. This rule allocates these total damages by going down a list of plaintiffs in the descending order of the probability that they were injured by defendant and fully compensating each such plaintiff until all the damages are paid.⁽⁸⁾

Such results have beguiled or bedeviled generations of legal scholars,⁽⁹⁾ economists,⁽¹⁰⁾ political scientists,⁽¹¹⁾ psychologists,⁽¹²⁾ philosophers,⁽¹³⁾ and decision theorists.⁽¹⁴⁾ They have made their way into evidence textbooks.⁽¹⁵⁾ They have been used by Supreme Court justices concerned with the constitutionality of relaxing the requirement of proof beyond a reasonable doubt⁽¹⁶⁾ and of diminishing the size of the criminal jury.⁽¹⁷⁾ They have emboldened courts to offer quantitative expressions for the burden of persuasion.⁽¹⁸⁾ They shed light on the vexing problem, left unresolved in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,⁽¹⁹⁾ of the relationship between statistical significance and legally sufficient or admissible statistical proof.⁽²⁰⁾ They have, in sum, shaped our understanding of the burden of persuasion.

Even so, mathematics cannot dictate legal policy any more than law can dictate mathematical truth.⁽²¹⁾ The application of technical concepts like probability, utility, and expected value to the legal process bears close scrutiny. This article discusses these concepts to clarify the assumptions and limitations of the decision-theoretic analyses. In doing so, it responds a proclamation of "the demise of legal theorems" about "evidential reasoning."⁽²²⁾ According to Professor Ronald J. Allen, a leading figure in the fields of evidence and procedure,⁽²³⁾ the "best example" of the death of "formalisms"⁽²⁴⁾ is "the various proofs that employing the civil burden of persuasion of a preponderance of the evidence will minimize or optimize errors."⁽²⁵⁾ Professor Allen believes that "[t]hey are all false as general proofs (although not as special cases), and all for the same reasons. They neglected base rates and the accuracy of probability assessments of liability"⁽²⁶⁾

Although Professor Allen's general reservations about taking "legal theorems" too seriously have obvious merit, his specific claims about the proofs fail to attend to a crucial distinction between expected and actual errors.⁽²⁷⁾ It is true that there are no general proofs as to what rules will minimize the occurrence of actual errors, but there are readily verifiable proofs, like those whose conclusions are stated above, that a given decision rule maximizes expected utility or minimizes expected losses. Finding the decision rule that minimizes the expected value of a prescribed loss function is an extremely general procedure, and the proofs remain true for all possible base rates.⁽²⁸⁾

This article presents one such proof and provides a few numerical examples that illustrate the sense in which the more-probable-than-not standard is optimal. This exercise clarifies both the premises and conclusions of the decision-theoretic analysis of the civil burden of persuasion. By describing the mathematical reasoning carefully, I hope to lay to rest common misconceptions about the properties of the rules and to indicate how the evidentiary analysis fits into a broader framework of economic and legal analysis. In short, I discuss both the mathematical reasoning and its implications for judges or code writers who must specify the burden of persuasion that a party to a lawsuit must carry and that a judge or jury charged with finding the facts must consider.

I. The Big Picture

According to the framework known as Bayesian decision theory, a "rational" actor will make those decisions that maximize subjective expected utility (or, equivalently, that minimize expected loss).⁽²⁹⁾ Working within this normative theory, many legal scholars have been impressed with its power to explicate the somewhat nebulous formulations of the burden of persuasion⁽³⁰⁾ in criminal and civil cases.⁽³¹⁾ The theory interprets phrases like "preponderance of the evidence" and "beyond a reasonable doubt"

as specifying decision rules in terms of a juror's subjective probability that the facts are such as to warrant imposing liability.⁽³²⁾ In particular, it interprets the preponderance standard to mean, "Return a verdict for plaintiff if the probability is greater than $\frac{1}{2}$ that the facts that plaintiff needs to prevail are as plaintiff alleges." The threshold probability that corresponds to the elimination of any "reasonable doubt" is much higher.

The difference in the criminal and civil burdens of persuasion and the transition point of $\frac{1}{2}$ in civil cases seem to flow naturally from the command to maximize expected utility (or minimize expected loss). In the simplest derivation, the transition point is just a function of the two error costs--the loss associated with a false verdict for plaintiff and the loss associated with a false verdict for defendant. When these two losses are of equal magnitude, the more-probable-than-not standard always minimizes the expected loss.⁽³³⁾ When they are different, a different threshold probability minimizes expected loss.⁽³⁴⁾ For example, if it is ten times worse to convict an innocent person charged with a crime than to acquit a guilty person, the threshold probability is no longer $\frac{1}{2}$, but $\frac{10}{11}$, or 0.91.⁽³⁵⁾

In bold, the Bayesian analysis of the usual civil burden of persuasion involves three essential claims: (1) that the decision rule should be one that minimizes expected loss; (2) that the loss whose expectation should be minimized is some quantity that has the same value when an erroneous verdict hurts a plaintiff as when an erroneous verdict hurts a defendant; and (3) that the more-probable-than-not rule always minimizes the expectation of that loss. These propositions all deserve scrutiny; some are more secure than others.

A. The Criterion: Minimize Expected Loss

For decades, the first proposition has proved controversial among legal scholars--and rightly so. There are criteria other than minimizing expected loss that seem appealing, at least at first blush.⁽³⁶⁾ Instead of being concerned just with the relative costs and utilities of erroneous verdicts for plaintiffs and erroneous verdicts for defendants, we might choose to pursue a more complex form of multi-attribute decisionmaking.⁽³⁷⁾ More radically, we might ignore the utilities entirely and seek a particular ratio in the actual or expected numbers of the two types of erroneous verdicts.⁽³⁸⁾ For instance, we might want to equalize the risk of an erroneous verdict for plaintiff and an erroneous verdict for defendant.⁽³⁹⁾

Although Bayesian decision theory--including its central commandment to maximize expected utility--rests on axioms about preferences for risk-related outcomes⁽⁴⁰⁾ that many writers find appealing,⁽⁴¹⁾ arguments about the acceptance of the axioms pervade the literature of philosophy, statistics, and economics.⁽⁴²⁾ It is next to impossible to convince determined skeptics by pointing to intuitively appealing formal axioms. Because there is less agreement on the suitability of the axioms of rational choice than there is on the three axioms of probability theory,⁽⁴³⁾ debate on the desirability of maximizing subjective expected utility continues.⁽⁴⁴⁾ In the hope of clarifying the debate in the field of law, however, I survey the main lines of argument in Part IV and sketch some reasons to believe that the goal of minimizing expected loss is as appropriate in law as it is in other decision problems.

B. The Definition of Loss: Symmetry as Between Plaintiff and Defendant

The second proposition--that the loss whose expectation should be minimized is some quantity that has

the same value when an erroneous verdict hurts a plaintiff as when an erroneous verdict hurts a defendant—has proved less controversial, although important questions can be raised. What, after all, do we mean by "loss"? If we are concerned merely with counting the correctness of verdicts, that is, with whether the jury gets it right or wrong regardless of the consequences to the litigants or to society, "loss" just means a wrong verdict. The notion that we should strive to avoid all types of errors with equal vigor is appealing if the legal system does not favor one type of litigant over the other. From the standpoint of a legislator seeking an appropriately general rule for deciding cases, the consequences of each type of mistaken verdict are the same. This symmetry implies, in the language of statistics, that the loss is a binary variable that takes on a fixed value (conventionally, zero) when the verdict is true and some other fixed quantity (conventionally, one) when it is false.

However, the losses associated with a false verdict in some cases may be different than the loss in others. An action to collect an overdue bill in small claims courts has different consequences than a class action in a securities fraud case. To account for this type of variation in the consequences of errors, we could take the loss to be monetary. For example, in a case in which (a) the only consequence of a false verdict for plaintiff is that the defendant pays out a given number of dollars to plaintiff that would not have been transferred in a world of perfect knowledge, and (b) the only consequence of a false verdict for defendant is that the plaintiff is deprived of the dollars that should have been awarded, we might define the loss as the money that stays in the wrong pockets.⁽⁴⁵⁾ Inasmuch as this definition also implies that, for any single case, the loss associated with each type of false verdict is the same, it preserves the symmetry between plaintiffs and defendants.

Nonetheless, there are other consequences to a verdict than just a transfer of dollars between the parties,⁽⁴⁶⁾ and there are costs associated with the process of litigation itself.⁽⁴⁷⁾ Furthermore, a central tenet of Bayesian decision theory is that the decisionmaker should minimize not simple monetary loss, but losses in utility.⁽⁴⁸⁾ That raises the question of whose utilities count and what they might be. Are they related to the preferences of the parties, the jurors, or some abstract, social entity?⁽⁴⁹⁾ If the objective is minimize the parties' expected losses in utility, then differences in risk-aversion between a plaintiff and a defendant would destroy the symmetry of the loss function.⁽⁵⁰⁾ Nevertheless, from the standpoint of designing a rule to be followed by juries, it seems more appropriate to think of the jurors as applying legislative or judicial preference-orderings rather than their own or those of the parties. If, from this external perspective, the consequences of mistaken verdicts, in whatever units they may be measured, are of the same magnitude for each side in a case, and if these are the only costs to be considered, then the second proposition in the derivation of the more-probable-than-not rule holds.⁽⁵¹⁾

C. From the Loss Function to the More-Probable-than-Not Rule

Until recently, the third proposition—that the goal of minimizing expected loss for a function that counts each type of error as equally bad leads to the more-probable-than-not rule in a two-party case—seemed beyond dispute.⁽⁵²⁾ More generally, there was little dissent from the claim that whether or not expected loss minimization is the appropriate objective, the traditional legal standards of civil and criminal cases follow from it. And if that is so, Bayesian decision analysis packs considerable power as a positive theory of this nook and cranny of the law of evidence.⁽⁵³⁾

For this reason, it is important to consider the third step in the decision theoretic analysis. Part II gives an exposition and informal proof of the proposition that Bayesian decision theory generates those decision

rules that minimize expected losses.⁽⁵⁴⁾ Part III analyzes putative counter-examples and shows that they do not relate to expected losses; hence, they do not undermine this result. Part IV examines some criteria other than the minimization of expected losses. It shows that arguments that Bayesian decision theory does not explain or justify the civil burden of persuasion because it does not imply that the more-probable-than-not standard will minimize the total number of errors (or equalize the actual numbers of each type of error) involve the first, and not the final step in the decision-theoretic program. As for that step, it indicates why the arguments to date have not unseated Bayesian decision theory and do not warrant modifying the more-probable-than-not rule in the usual civil case.

II. Minimizing Expected Losses

Much of the apparent disagreement about the application of Bayesian decision theory to law relates to the use of technical terms like "expected value." To clarify the meaning of the crucial terms, and to show the connection between Bayesian decision theory and a burden of persuasion--without preconceptions about the law--let us consider a stylized but everyday example of a decision problem. I am about to walk from my home to my office.⁽⁵⁵⁾ Should I take my umbrella with me?

Bayesian decision theory offers a way to decide. Let R be a random variable that describes whether it will rain during my walk. Thus, R can take on only two values (denoted by r); let $r=0$ indicate that it will not rain, and let $r=1$ indicate that it will. The "action space" consists of two points: d_0 (leave the umbrella at home), and d_1 (take the umbrella).⁽⁵⁶⁾ The loss function describes the adverse consequences L_{10} , of taking the umbrella when it does not rain ($d_1, r=0$), and L_{01} , of not taking the umbrella when it does ($d_0, r=1$).⁽⁵⁷⁾ That is, L_{10} represents the cost of the wasted effort in taking the umbrella; L_{01} is the cost of getting wet (or ducking in somewhere out of the rain).

I could decide between d_0 and d_1 arbitrarily, say by flipping a coin, or by always taking the umbrella. But I can do better if I know the probability of rain. Toward this end, I can gather data. If I look out the window, I might see clouds. I might even see rain.⁽⁵⁸⁾ Or, I might obtain a professional forecast for the region. Based on the best available data, I do my best to evaluate the probability of rain $p_1 = p(r=1 | \text{data})$ and its complement, $p_0 = p(r=0 | \text{data})$.⁽⁵⁹⁾ It will be convenient to write these probabilities as p and $1-p$, respectively.

Armed with these subjective but data-influenced probabilities, what should I do? That depends on what I want to accomplish. I cannot decide on the basis of what actually happens because I do not yet know what will happen. In other words, I cannot minimize *actual* loss. But I can use the probabilities to find the *expected* loss and minimize it.⁽⁶⁰⁾

The expected value of a variable is the weighted average of its possible values, where the weights are the probabilities of these values. For example, if a pair of dice are thrown, the total number of spots is a random variable. The chance of two spots is $1/36$, the chance of three spots is $2/36$, and so forth; the most likely number is 7, with chance $6/36$. The expected value is therefore

$$(2 \times 1/36) + (3 \times 2/36) + (4 \times 3/36) + (5 \times 4/36) + (6 \times 5/36) + (7 \times 6/36) + (8 \times 5/36) + (9 \times 4/36) + (10 \times 3/36) + (11 \times 2/36) + (12 \times 1/36) = 7$$

Likewise, the expected loss of d_0 (leaving the umbrella) is $L_{01} \times p_1 = L_{01} \times p$, and the expected loss of d_1 is $L_{10} \times p_0 = L_{10} \times (1 - p)$. The former quantity is the cost of leaving the umbrella when it rains, discounted by the probability that it will rain; the latter is the cost of taking the umbrella, weighted by the probability that it will not be needed. These results are summarized in Table 1.

	<i>States of Nature</i>	
	Dry	Rain
<i>Acts</i>	<i>Probability</i>	
	$1 - p$	p
Leave umbrella (d_0)	0	pL_{01}
Take umbrella (d_1)	$(1 - p)L_{10}$	0

Table 1. Expected losses for the two acts

Suppose I never take the umbrella. As shown in Figure 1, my expected loss under this rule is an ascending straight line (as a function of p) with slope L_{01} . This simple rule makes no actual errors when it does not rain, but even on those days, there is an expected loss of pL_{01} . The rule works well when the probability of rain is low (small p), but the expected loss grows as the chance of rain increases.

Now consider the decision rule that instructs me always to take the umbrella. This rigid rule makes no actual errors when it does rain, but even on those days, it gives an expected loss of $(1 - p)L_{10}$. This expected loss produces the downward sloping straight line in Figure 1. The take-the-umbrella rule works well when the probability of rain is high (large p , small $1 - p$), but it performs badly when the chance of rain is low.

Clearly, the best approach is to leave the umbrella when the chance of rain is low and to take it when it is high. The trick is to find the precise point to switch from the leave-the-umbrella to the take-the-umbrella rule. Figure 1 reveals that this point occurs when (or just after) the expected-loss line for taking the umbrella intersects the expected-loss line for leaving it. By switching at this point, we follow the line with the smaller expected loss.

But precisely where does the point of intersection occur? The lines intersect at the value of p where their heights are equal. Designating the value of p at this point as p^* , we simply solve the equation $p^*L_{01} = (1 - p^*)L_{10}$ to arrive at $p^* = L_{10} / (L_{01} + L_{10})$. Switching from one decision to the other at p^* gives the expected loss displayed in Figure 1 as the darkened upper two sides of the triangle with a vertex perpendicularly above the transition value p^* .

Figure 1. Expected loss as a function of the probability of rain p if I take the umbrella (d_1) and if I do not (d_0). I am indifferent at the transition probability $p^* = L_{10} / (L_{01} + L_{10})$, and an optimal decision rule is to switch from d_0 to d_1 when $p > p^*$.

In sum, to minimize my expected loss, an optimal decision rule is to take the umbrella (d_1) as long as the expected loss of leaving it exceeds the expected loss of taking it, and to leave it (d_0) otherwise. That is, I choose d_1 if and only $p_1 \times L_{01} > p_0 \times L_{10}$, which is to say:

choose d_0 if $p(r=1) L_{10} / (L_{01} + L_{10})$; choose d_1 if $p(r=1) > L_{10} / (L_{01} + L_{10})$. (1)

Now, this result holds for *any* values of the losses L_{10} and L_{01} . In Figure 1, L_{01} is a little larger than L_{10} , but if I were ten times as anxious to avoid carrying the umbrella unnecessarily as to avoid being caught in the rain without an umbrella ($L_{10} = 10 \times L_{01}$), the vertex would move to the right, and p^* would equal $10 / (10 + 1) = 10/11$; I would carry the umbrella only if I judged the chance of rain to exceed 10/11. Figure 2 shows this situation.

Figure 2. An expected-loss minimizing decision rule when $L_{10} = 10 \times L_{01}$ is to switch from d_0 to d_1 when $p > 10/11$.

If I were indifferent as between taking the umbrella unnecessarily and leaving it when it turned out to be needed ($L_{10} = L_{01}$), the transition probability would be

$$p^* = 1/(1 + 1) = 1/2.$$

The optimal decision rule then is to take the umbrella as long as the probability seems to favor rain.

The analogy to the more-probable-than-not standard of civil litigation is close at hand. The decision to take the umbrella is like a plaintiff's verdict. Rain is like the state of affairs that plaintiff alleges and that would lead to recovery under the applicable law. L_{10} is the social cost of a mistaken verdict for plaintiff; L_{01} is the social cost of a mistaken verdict for defendant. If these costs are equal, then the triangle becomes isosceles, and the transition value becomes $p^* = L_{10} / (L_{10} + L_{01}) = 1/2$. The jury should return a plaintiff's verdict whenever the probability of the facts that establish plaintiff's claim exceeds 1/2. All this is depicted in Figure 3.

Figure 3. Expected loss as a function of the probability of liability for a plaintiff's verdict (d_1) and a defendant's verdict (d_0) when $L_{01} = L_{10}$. The transition probability is $p^* = L_{10} / (L_{01} + L_{10}) = 1/2$, and an optimal decision rule is to return a verdict for plaintiff only when $p > 1/2$.

Furthermore, the optimal decision rule minimizes the expected losses for any and all values of the probability p . Because the rule works for each value of p , it works no matter how p happens to be distributed in any batch of actual cases. Consequently, applying it to the umbrella decision on all days (and analogously, to reach a legal verdict across all cases) minimizes the total expected losses over time. There is no need to consider the frequency of apparently or actually meritorious cases to minimize expected loss in each case.

Finally, although the binary decision problem is sufficient to handle the major situations of legal interest, still more general methods to arrive at an optimal decision rule--one that always minimizes expected losses--can be applied to handle situations with more than two decisions and more than two states of nature.⁽⁶¹⁾ For all these reasons, it seems to fair to conclude that finding the decision rule that minimizes the expected value of a prescribed loss function is an extremely general procedure.

III. Professor Allen's Attack

Some commentators are dubious of this conclusion. Although arriving at an optimal decision rule flows ineluctably from the definition of expected value and the rules of algebra, Professor Allen writes:

"Prof. Kaye asserts that no matter what the base rate, his theory of expected losses applies equally well, and that it has nothing to do with the number of errors, so long as 'every erroneous verdict for a plaintiff entails the same loss as every erroneous verdict for a defendant.' If this were true, it would be astonishing. On the basis of very little substantive knowledge--all you know is a little algebra and that 'every erroneous verdict for a plaintiff entails the same loss as every erroneous verdict for a defendant'-- a general

decision making algorithm appears that will maximize your expected utility, and it has nothing to do with error minimization, as I in a burst of silliness, suggested. Really?⁽⁶²⁾"

As we have just see, the optimal decision rule in expression (1) for the umbrella problem does not involve any "base rates," but only subjective probability and relative losses. Of course, my probability judgment may well be influenced by the knowledge that I have of the "base rate" for rain, but the optimal decision rule is the same whether p is formulated with or without such knowledge.⁽⁶³⁾

Professor Allen continues:

Compare two worlds, one in which there are 100 errors and one in which there are 101. In which world, in Prof. Kaye's terms, would we have a greater expected loss? Remember that we know nothing about the actual distribution of errors or their size, because Prof. Kaye's world is one largely devoid of substantive knowledge. Obviously we would have greater expected loss in the world with 101 errors.⁽⁶⁴⁾

Several mistakes are apparent here. First, the decision rule cannot be applied in a world "devoid of substantive knowledge." Quite the contrary, the rule exploits our knowledge of the world to arrive at the probabilities of the states of nature.⁽⁶⁵⁾ Deprived of all knowledge of the world, I would be hard pressed to gauge the probability of rain. Living in the real world, like any juror, I can make such judgments.

Second, if at the end of the year, I know that my umbrella decisions generated a specific number of errors, I also know the days on which I erred. At the beginning of each day, I had only ex ante knowledge, and that required me to resort to probabilities. At the end of each day, I know more.

Third, one cannot just assume that the expected loss is greater in World-101 than in World-100.⁽⁶⁶⁾ Taking the losses to be of equal magnitude ($L_{01} = L_{10} = 1$), suppose that in World-100, where there were 100 wrong decisions, the probability of rain was $p = 0.1$ for the first fifth of the year, 0.4 for the next fifth, and so on, as shown in column 2 of Table 2. The 100 actual errors are shown in column 3. The expected losses are computed according to the optimal switching rule shown in Figure 2. They total 102.2 in column 4.

Period	p	Actual loss	Expected loss
1 st fifth	0.1	7	7.3
2 nd fifth	0.4	22	29.2
3 rd fifth	0.5	37	36.5
4 th fifth	0.7	22	21.9
5 th fifth	0.9	12	7.3
Total		100	102.2

Table 3. Hypothetical expected and actual numbers of errors with a transition probability $p = 1/2$. The expected loss depends only on the numbers of days at each probability value.

Now for World-101. The numbers in Table 4 produce a total actual loss of 101, but the same expected

loss of 102.2.

Period	p	Actual loss	Expected loss
1 st fifth	0.2	15	14.6
2 nd fifth	0.3	22	21.9
3 rd fifth	0.5	37	36.5
4 th fifth	0.7	22	21.9
5 th fifth	0.9	5	7.3
Total		101	102.2

Table 4 Different hypothetical expected and actual numbers of errors with the $p > \frac{1}{2}$ rule. As in table 3, expected loss depends on the numbers of days at probability p and not the actual numbers of errors.

Similar tables could be produced to show not only that the expected losses could have been identical in each hypothetical world, but that they could have been larger in the second than the first, or that they could have been smaller.

Of course, I do not mean to say that there is no connection whatever between the observed values of a random variable and its expected values.⁽⁶⁷⁾ Generally, the observed values will approximate the expected values, so there is a statistical, though not a necessary connection, between expected losses and actual losses. The statistical law of large numbers and its corollaries, however, do not undermine the point that Bayesian decision theory offers a general method of finding the decision rules that minimize expected losses.⁽⁶⁸⁾

With these matters clarified, we arrive at the nub of Professor Allen's argument:

"Even more remarkable is Prof. Kaye's assertion that his 'proofs remain true for all possible base rates.' Remember what the proof is--it is a proof that a certain rule, preponderance of the evidence, will minimize expected losses. I asserted it is true in only a limited number of situations. He says 'The proofs remains true for all possible base rates.' We have already established that we use words differently, so perhaps I misunderstand what 'true' means. Let me be clear why I think this is false. Consider a world in which no deserving defendants go to trial, and for some deserving plaintiffs the fact finder assesses the likelihood of their case [sic] to be .5 or less. All such cases are errors, offset by no competing errors for the defendant. In this world, is the .5 rule 'optimal'? Obviously not. Lowering the standard can only reduce the total number of errors and thus the total expected loss (although one would have to worry about secondary consequences). Thus, the assumptions underlying Kaye's proof turn out to be quite rigorous. The base rates and the assignments of probabilities have to be in particular relationships in order for any rule to minimize expected losses. In the infinite number of worlds in which these relationships do not hold, expected losses will not be minimized."⁽⁶⁹⁾

Professor Allen is offering a numerical counter-example to refute the general proof given in Part II.⁽⁷⁰⁾ Unless there is some mistake in the proof itself, this project is doomed to failure. Nevertheless, to examine the claim more directly, a more complete putative counter-example of this type is given in Table

5, in the context of the umbrella decision. It posits a most unusual year. The best data I had on the chance of rain each day—including my knowledge of the incidence of rain for all previous years—suggested that there would be about 190 rainy days. Yet, there were 365. I was not wrong to leave the umbrella when the odds did not favor rain: unexpected floods happen. As a result, the $p > \frac{1}{2}$ rule led me to make 74 reasonable, but ultimately mistaken judgments about when to carry the umbrella. Still, as in Table 4, I kept the *expected* losses to the minimum possible value of 102.2.

Period	p	Actual loss	Expected loss	Expected rainy days
1 st fifth	0.2	15	14.6	14.6
2 nd fifth	0.3	22	21.9	21.9
3 rd fifth	0.5	37	36.5	36.5
4 th fifth	0.7	0	21.9	51.1
5 th fifth	0.9	0	7.3	65.7
Total		74	102.2	189.8

Table 5. Hypothetical expected and actual numbers of errors with the $p > \frac{1}{2}$ rule

Could I have done better—in terms of expected loss? Is it true that “[l]owering the standard can only reduce the total number of errors and thus the total expected loss”?⁽⁷¹⁾ Let us drop the transition probability from $\frac{1}{2}$ to $\frac{1}{4}$. Table 6 shows that the expected loss under this decision rule increases from 102.2 to 131.4. The modified decision rule reduces the actual loss from 74 to 15, but it does not and cannot reduce the expected loss.

Period	p	Actual loss	Expected loss
1 st fifth	0.2	15	14.6
2 nd fifth	0.3	0	51.1
3 rd fifth	0.5	0	36.5
4 th fifth	0.7	0	21.9
5 th fifth	0.9	0	7.3
Total		15	131.4

Table 6. Hypothetical expected and actual numbers of errors with a $p > \frac{1}{4}$ decision rule. In this example, that rule decreases the actual loss compared to an optimal rule, but it increases the expected loss. In general, departing from the $p > \frac{1}{2}$ rule increases expected loss and might or might not decrease actual loss.

Now, one is free to argue that using a number like $\frac{1}{4}$ instead of $\frac{1}{2}$ for the transition probability is better at minimizing *actual* errors,⁽⁷²⁾ but this strategy cannot slay the dragon of Bayesian decision in its own lair of expected losses. Rhetoric is no match for arithmetic,⁽⁷³⁾ and the final step in the decision theoretic program can be taken with confidence.

But what of the first step? Why should we worry about expected loss when litigants must live with actual errors? Having clarified the mathematics of minimizing expected loss, it is time to examine the implications for the legal system and for actual rather than expected errors.

IV. Implications for the Law

Although the mathematics of minimizing expected value is straightforward, the translation of the mathematical results into explanations of or prescriptions for legal procedures is neither automatic nor trivial. Various questions arise: What types of knowledge are necessary to derive and apply an optimal decision rule? What if jurors' personal probabilities are internally coherent but epistemologically foolish? Can or should the legal system search for and implement some rule that would minimize the incidence of actual errors or achieve some particular mix of errors rather than merely minimizing expected errors? This section considers such questions.

A. Substantive Knowledge

Because the analysis that leads to the $p > \frac{1}{2}$ rule seems to attend to form rather than substance, it might be thought that the rule is entirely mechanical in its derivation or application.⁽⁷⁴⁾ However, substantive knowledge is involved in the derivation of the decision rule and especially in the application of that rule. With respect to the derivation, knowledge of the structure and goals of the legal system is essential to ascertaining the losses that determine the transition probability p^* . Of course, this knowledge is rather thin in that the derivation of the $p > \frac{1}{2}$ rule requires no knowledge of any substantive domain beyond that required to specify the loss function. This feature, however, is a strength rather than a weakness of the analysis. It reflects the powerful generality of a theory that can be applied to any decision problem--from carrying an umbrella when it is cloudy, to drilling for oil, to investing in the stock market, to selecting the best therapeutic regimen for a sick patient, to deciding whether the disputed facts in a lawsuit are such as to give rise to liability.

Second, we need considerably more substantive knowledge to *apply* a decision rule that minimizes expected loss. The decision rule follows logically and formally from the theory, but it comes into play only *after* informal, substantive, and practical knowledge has been used to arrive at the probabilities on which the decision turns in the fashioned prescribed by the rule.⁽⁷⁵⁾

B. Representation Theorems

Although using Bayesian decision theory to arrive at a decision rule for adjudication is not a mechanical task, and although the decision rule cannot be applied according to any known algorithm, a deeper issue must be confronted before the analysis can explain or justify the more-probable-than-not standard. Why should a decisionmaker strive to minimize expected loss in the first place?

One possible justification looks to the long run properties of decision rules. When averaged over many cases, actual results tend to converge on expected values. For instance, suppose that a poll of a simple random sample of the voting-age population shows that 67% favor campaign finance reform. One argument for the common practice of taking 67% as an estimate of the proportion in the entire population is that the expected value of the sample proportion is the population proportion. That is, if repeated random samples were taken, some sample proportions would exceed the population value (whatever it might be), some would be less than that value, and a very few might be right on target.

Averaged over all possible samples weighted by the probabilities of their occurrences, the sample proportions equal the population proportion.⁽⁷⁶⁾

When the same gamble must be taken repeatedly, this rationale for the expected value criterion can be convincing, but it has much less force when the gamble is taken only once. Because each case that comes to trial is unique, the long-run argument seems insufficient. Probabilists have responded to the problem of defining probabilities and decision rules for unique events with theories that start with a rich set of personal preferences with respect to the outcomes of gambles. These axiomatic theories of rational choice imply that the appropriate strategy remains the maximization of expected utility.

Legal scholarship has not probed very deeply into these theories. Claims that the axioms are peculiarly inapposite to legal factfinding typically suffer from a failure to consider the axioms themselves and the justifications that have been offered for them.⁽⁷⁷⁾ It seems appropriate, therefore, to pause to convey something of the flavor of the axiomatic theories. It must be emphasized that in these theories, attributions of probability and utility are not fundamental; rather, they are merely a way of conceptualizing preferences for various outcomes (that are not necessarily sure to occur), like being rained on or carrying an umbrella on a day that may well be sunny.⁽⁷⁸⁾ So-called representation theorems show that if a person's preferences satisfy certain qualitative conditions, then those preferences can be represented as maximizing expected utility relative to some probability and utility functions.⁽⁷⁹⁾

There are many such theorems, each making somewhat different assumptions. One famous formulation comes from the statistician Leonard J. Savage.⁽⁸⁰⁾ Professor Savage defined a "weak preference" for an act g over an act f to mean preferring g to f or else being indifferent between them. Using the notation $f \text{ wpt } g$ to indicate that g is weakly preferred to f , for any acts f , g , and h , Savage's first postulate is that wpt is a "simple ordering" in that the following two conditions hold:

connectedness: either $f \text{ wpt } g$ or $g \text{ wpt } f$ (or both), and

transitivity: if $f \text{ wpt } g$ and $g \text{ wpt } h$, then $f \text{ wpt } h$.

Savage's second postulate asserts a property of *independence*: if two acts have the same consequences in some states, then preferences regarding those acts should be independent of what the common occurrence is.⁽⁸¹⁾ Savage shows that if preferences satisfy these and other assumptions,⁽⁸²⁾ then they maximize expected utility relative to some probability and utility functions.

At this level, the argument against Bayesian decision theory must be that the preferences we want jurors to implement lack such properties as connectedness, transitivity, or independence.⁽⁸³⁾ Some legal writers have suggested that we simply do not care about deciding in accordance with appropriately ordered preferences because real jurors cannot satisfy the axiomatic constraints due to limitations of time, resources, or processing power.⁽⁸⁴⁾ Clearly, actual jurors may not arrive at precisely the same value for their subjective probabilities as a Bayesian juror with infinite time and computational resources. Likewise, an elementary school student asked to divide 153,387 by 513 in 10 seconds might not arrive at precisely the same value for the quotient as a student with pencil, paper, the algorithm for long division, and ten minutes to work out the answer. Nevertheless, we could ask the first student to use the algorithm to divide 150,000 by 500 and thereby obtain an excellent approximation. Simplifications are inevitable and important in all applications of decision theory to realistic problems.⁽⁸⁵⁾ The law is no exception.⁽⁸⁶⁾

Practical and procedural constraints might well suggest that jurors should not be asked to assimilate each item of evidence in quantitative terms--a conclusion that is itself consistent with Bayesian decision theory⁽⁸⁷⁾--but they do not necessarily mean that we should abandon the effort to minimize expected loss and instruct jurors that they may as well return a verdict for the party on the basis of an apparently improbable set of facts, or that they should return a verdict for defendant even though the odds are such that plaintiff deserves to prevail.⁽⁸⁸⁾

C. Adjusting for Juror Error

Although resource constraints cloud but do not remove the appeal of the expected utility principle, the possibility that jurors will systematically misjudge probabilities must be considered before applying the principle.⁽⁸⁹⁾ Suppose, for example, we somehow knew that jurors in civil cases typically misjudge probabilities in favor of plaintiffs by 0.1. (In statistics, such systematic as opposed to random error would be called bias.) We might respond by using

$$p^* = 0.1 + L_{10} / (L_{01} + L_{10})$$

as the threshold value for a plaintiff's verdict (instead of the usual value $L_{10} / (L_{01} + L_{10})$).⁽⁹⁰⁾ This is much like aiming a target pistol a few degrees off center to compensate for a distortion in the muzzle. We would not be abandoning the criterion of hitting the target (minimizing expected loss),⁽⁹¹⁾ but improving our implementation of it.⁽⁹²⁾

Moreover, the need to correct for bias in probability judgments is hardly restricted to the legal realm. I enjoy climbing mountains, but I worry that I tend to underestimate the level of risk. If I were to perform expected value calculations to decide whether to climb a particular peak, I might want to compensate for my bias by reminding myself to think very carefully about all the dangers of the route, or simply by adjusting my sincere estimate of the risk upward; it is less likely, but still possible that I might want to use an unadjusted subjective value for that risk together with a modified decision rule that would require additional expected utility to warrant the climb. Both approaches adjust for a known bias.

Justifying such modifications of the optimal decision rule (1) for verdicts, however, would be extraordinarily difficult.⁽⁹³⁾ It is far from obvious that jurors' probability estimates in civil cases (or in many identifiable subsets of civil cases) ordinarily are skewed in only one direction.⁽⁹⁴⁾ Furthermore, even if we knew that jurors consistently overestimate the strength of one side's case, the solution might not lie in altering the decision rule, but in mitigating the factors that prompt such overestimates. Rather than aiming the pistol off-center, it might be better to fix the muzzle. The rules of evidence, for instance, warrant the exclusion of evidence whose probative value is likely to be misjudged;⁽⁹⁵⁾ if jurors routinely are overimpressed by gory photographs or videos of the victims of tragic accidents, the solution lies in limiting such presentations, not in ratcheting the burden of persuasion up an arbitrary notch.

D. Actual Errors

The decision rule (1) of Bayesian decision theory always minimizes expected loss (relative to the decisionmaker's utility and probability functions). However, some writers have suggested that it would be better to "minimize" or "optimize" the actual number of errors.⁽⁹⁶⁾ Identifying a decision rule with this

property and devising a corresponding jury instruction about the burden of persuasion would be a nightmare.⁽⁹⁷⁾ Suppose that 30% of all cases are meritorious in the sense that if only we knew the true state of affairs, the applicable legal rules would dictate liability. Should we raise the burden of persuasion by requiring a threshold of $p = 0.7$ for a plaintiff's verdict? Should we stick with the usual 0.5? To analyze the effects of such decision rules on the numbers of correct and incorrect verdicts, we would need to know not only the percentage of meritorious cases, but also how the subjective probability p is distributed across both non-meritorious and meritorious cases.⁽⁹⁸⁾

Figure 4 shows how this works, at least in principle. The non-meritorious cases lie in a triangular region to the left of (but overlapping) the meritorious cases, which, on average, have higher apparent probabilities of liability.⁽⁹⁹⁾ The curves f_0 and f_1 for these two types of cases are "densities"; the total area under each curve indicates the number of cases of each type.⁽¹⁰⁰⁾ If only 30% of all cases were meritorious, the area under f_1 would be 30% of the area under both curves. Thus, the base rates are reflected in the areas. The transition point for verdicts of liability that minimizes the total expected loss is p^* . Non-meritorious cases to the right of p^* are false verdicts of liability. Their number is indicated by the hatched area to the right of p^* and under f_0 . Meritorious cases to the left of p^* are false verdicts of non-liability. Their number is indicated by the hatched area to the left of p^* and under f_1 . Hence, the number of actual errors is the total hatched area. The transition probability that would minimize this combined area is not necessarily p^* . In Figure 4, we could do better by moving to a higher transition probability; as the dashed line moves to the right, the shaded area below f_0 and above f_1 disappears.

Figure 4. Searching for a transition probability that minimizes the actual, total number of errors. That number is the sum of the two hatched areas. Using a transition probability to the right of p^* would reduce actual error.

A major problem with using a picture like this to find the rule that minimizes the actual number of errors is that the curves in Figure 4 are fantasies. Because we will never have this kind of information about cases that go to trial, we have no choice but to ask the jury to use a threshold probability p^* that minimizes the expected number of errors but that cannot be guaranteed to minimize the actual number.

Nonetheless, the recognition that minimizing the incidence of actual errors is beyond our power does not mean that the more modest goal of at least reducing their incidence is beyond our reach. The use of the more-probable-than-not standard is but one of many legal policies or procedures designed to lower the risk of factually erroneous verdicts. As I have emphasized, the more-probable-than-not rule in the two-party civil case minimizes the expected number of erroneous verdicts, and it has the advantage of doing so whether the percentage of meritorious claims is 0%, 100%, or anything in between. The $p > \frac{1}{2}$ rule may not produce the minimum number of actual errors in any finite time period, but it is hard to know what rule would do better. Thus, to the extent that minimizing the expected number of erroneous verdicts tends to lower the actual number, policymakers who reject the fundamental tenets of BDT but still want reduce the risk of error should find it appealing.

The conditions under which minimizing the number of expected errors tends to reduce the actual number can be shown, in non-Bayesian terms, with an example. Suppose that some very large number N of civil cases are tried. As a case is decided, the verdict is placed in one of two piles--the jury-finds-liability and the jury-finds-no-liability pile. Furthermore, the jury's estimate P of the probability of liability is written on the verdict form. Now assume that the juries are perfectly well calibrated.⁽¹⁰¹⁾ That is, of every n_i cases determined to have probability p_i , exactly $p_i n_i$ belong in the liability pile.⁽¹⁰²⁾ Since there is no way to differentiate further these n_i cases in terms of the chance that they truly belong in the liability pile as opposed to the non-liability pile, the juries must make their best guesses as to all of them. Under the more-probable-than-not standard, all n_i go into the liability pile when $p_i > \frac{1}{2}$, and into the non-liability pile otherwise.

As a long-run strategy, this really is the best that we can do. If we were take a case from the liability pile and put it in the non-liability pile, the odds are that we would be making a mistake, for more than half those cases belong where they are, and we cannot tell one from the other. To nail this point down, consider the effect of moving 1,000 cases marked with probability-of-liability $p_i = 0.7$ from the liability to the non-liability pile. Subject to sampling error, $1000p_i = 700$ of the moves would be mistakes compared to $1000(1 - p_i) = 300$ correct moves. Of course, we could get lucky; it is theoretically possible that even though the expected outcome of the transfers is an increase of $1000(2p_i - 1) = 400$ errors, we happened to choose a sample of cases in which the majority were such that the true facts did not justify liability. But this is unlikely; and as we approach moving all n_i verdicts from the liability to the non-liability pile, we are certain to increase the number of misclassifications by $n_i(2p_i - 1)$.⁽¹⁰³⁾ In sum, the expected-error-minimizing rule administered by well calibrated juries tends to minimize actual errors.

If we relax the assumption that juries are perfectly calibrated, however, the distribution of the probability of liability over a set of cases can make a difference. Some verdict forms marked with the probability of liability are mislabeled--they should have some other value marked on them. If these discrepancies are large enough to move a case from one pile to another and if they move cases predominantly in one direction, then minimizing expected error over the estimated rather than the true probability values might not tend to minimize the actual errors.⁽¹⁰⁴⁾

Again, however, modifying the decision rule seems most unpromising as a method of reducing the actual number of erroneous verdicts. If the probability-estimation errors, on average, favored plaintiffs (or defendants), and if we knew the amount of that bias, then we could correct for it by tinkering with the

decision rule. However, there is no obvious reason to believe that the juries' errors in estimating p_i are biased,⁽¹⁰⁵⁾ and if the probability-estimation errors are unbiased, then the average long-run result of adhering to the $p > \frac{1}{2}$ rule will converge to that for well calibrated juries. With no knowledge of the degree of bias, or even whether any exists, it is difficult to imagine what instruction a jury should be given other than the conventional one that promotes actual verdict-error minimization with accurately ascertained probabilities. Under these generally applicable conditions, the more appropriate strategy for legislators concerned with minimizing actual verdict errors is to implement evidentiary and procedural rules that enhance the ability of juries to ascertain p accurately and without bias.

Conclusions

The decision rule (1) that follows from Bayesian decision theory necessarily minimizes expected loss--regardless of the "base rates" for true and false claims, and even though it is always possible that the judge or jury will not use the best possible value for the probability that the facts are such that liability attaches. Reports of the falsity and death of these results are greatly exaggerated. On the other hand, if one rejects the framework of Bayesian decision theory, but tries to minimize the actual number of verdict errors, no general solution is available. Realistically, however, the most plausible prescription to achieve actual error minimization still seems to involve instructing jurors to use the more-probable-than-not standard in a simple civil case. This decision rule minimizes expected errors (and, in the long-run, actual errors) with well calibrated juries.⁽¹⁰⁶⁾ To improve the calibration of actual juries, the rules of evidence and procedure should be structured to encourage the production and presentation of evidence in a manner that elucidates the probability of the true state of affairs on which liability depends.

These conclusions are not surprising. No mathematical result is self-applying, and additional argument is necessary to bridge the gap from a general mathematical truth to a substantive application--in law as in any other domain. I write to ensure that criticisms of Bayesian decision theory in understanding and justifying the law's burdens of persuasion be based on the theory as it exists and has been used. I do not claim that Bayesian decision theory is the only way to understand the burden of persuasion. Neither do I insist that decision rules that do not minimize expected loss (or maximize expected utility) might not somehow serve the law better. But no one has made a case for such standards or explained how they could be implemented, and no one has constructed a more revealing explanation of the law as it stands than that which flows from Bayesian decision theory.⁽¹⁰⁷⁾

NOTES

1. Regents' Professor, Arizona State University College of Law. B.S., MIT; M.A., Harvard University; J.D., Yale Law School. Ronald Allen and Richard Friedman generously provided comments on a draft of this paper, and Michael DeKay offered exceptionally helpful advice and corrections on a number of points. [\[BACK\]](#)
2. John Allen Poulos, *Innumeracy: Mathematical Illiteracy and Its Consequences* 134 (1988). [\[BACK\]](#)
3. *Grayned v. City of Rockford*, 408 U.S. 104, 110 (1972). [\[BACK\]](#)
4. See, e.g., John Kaplan, *Decision Theory and the Factfinding Process*, 20 Stan. L. Rev. 1065, 1071-72 (1968); Richard Lempert, *Modeling Relevance*, 75 Mich. L. Rev. 1021, 1032-35 (1977); Laurence Tribe, *Trial by*

Mathematics: Precision and Ritual in the Legal Process, 84 Harv. L. Rev. 1329, 1378-81 (1971). [\[BACK\]](#)

5. See, e.g., Kaplan, *supra* note 3; Lempert, *supra* note 3. [\[BACK\]](#)

6. See David Kaye, *Naked Statistical Evidence*, 89 Yale L.J. 601 (1980) (book review). [\[BACK\]](#)

7. D.H. Kaye, *The Limits of the Preponderance of the Evidence Standard: Justifiable Naked Statistical Evidence and Multiple Causation*, 1982 Am. Bar Foundation Research J. [now J. L. & Soc. Inquiry] 487, reprinted in *Evidence and Proof* (William Twining & Alex Stein eds., 1992). Professors Ronald Allen, Richard Kuhns, and Eleanor Swift credit me with having "demonstrated algebraically, if certain conditions are met the preponderance of the evidence standard should result in about the same number of errors being made for plaintiffs as for defendants." Ronald J. Allen et al., *Evidence: Text, Cases, and Problems* 828 (2d ed. 1997). That is not what that paper (or any other that I have written) shows. [\[BACK\]](#)

8. Daniel Farber, *Toxic Causation*, 71 Minn. L. Rev. 1219 (1987). The formal proof that this "most likely victim" rule minimizes the expected number of dollars left in the wrong pockets given in the appendix to this article is not correct; however, the proposed rule does minimize expected error defined in this fashion under the idealized conditions stated in the article. [\[BACK\]](#)

9. E.g., Vaughn C. Ball, *The Moment of Truth: Probability Theory and Standards of Proof*, 14 Vand. L. Rev. 807 (1961), reprinted in *Essays on Procedure and Evidence* 84 (1961); Robert Birmingham, *Remarks on "Probability" in Law: Mostly, a Casenote and a Book Review*, 12 Ga. L. Rev. 535 (1978); Richard S. Bell, *Decision Theory and Due Process: A Critique of the Supreme Court's Lawmaking for Burdens of Proof*, 78 J. Crim. L. & Criminology 557 (1987); James Brook, *Inevitable Errors: The Preponderance of the Evidence Standard in Civil Litigation*, 18 Tulsa L.J. 79 (1982); Alan D. Cullison, *Probability Analysis of Judicial Fact-Finding: A Preliminary Outline of the Subjective Approach*, 1969 Toledo L. Rev. 538; David Hamer, *Civil Standard of Proof Uncertainty: Probability, Belief and Justice*, 16 U. Sydney L. Rev. 506 (1994); D.H. Kaye, *Apples and Oranges: Confidence Coefficients Versus the Burden of Persuasion*, 73 Cornell L. Rev. 54 (1987); Lempert, *supra* note 3; Saul Levmore, *Probabilistic Recoveries, Restitution, and Recurring Wrongs*, 19 J. Legal Stud. 691, 696 n.8 (1990); Lawrence B. Solum, *You Prove It! Why Should I?*, 17 Harv. J.L. & Pub. Pol'y 691 (1994); Alan L. Tyree, *Proof and Probability in the Anglo-American Legal System*, 23 Jurimetrics J. 89 (1982); Tribe, *supra* note 3; cf. Kate Stith, *The Risk of Legal Error in Criminal Cases: Some Consequences of the Asymmetry in the Right to Appeal*, 57 U. Chi. L. Rev. 1 (1990). [\[BACK\]](#)

10. Terry Connolly, *Decision Theory, Reasonable Doubt, and the Utility of Erroneous Acquittals*, 11 L. & Hum. Behav. 101 (1987); Jason S. Johnston, *Bayesian Fact-finding and Efficiency: Toward an Economic Theory of Liability Under Uncertainty*, 61 So. Cal. L. Rev. 137 (1987); Richard Posner, *An Economic Approach to Legal Procedure and Judicial Administration*, 2 J. Legal Stud. 399 (1973). [\[BACK\]](#)

11. Bernard Grofman, *Mathematical Models of Juror and Jury Decision-Making: The State of the Art*, in *The Trial Process* 305 (Bruce D. Sales ed., 1981); Stuart Nagel, *Bringing the Values of Jurors in Line with the Law*, 63 *Judicature* 189 (1979); *The Model of Rules and the Logic of Decision*, in *Modelling the Criminal Justice System* 225 (Stuart S. Nagel ed., 1971); Stuart Nagel et al., *Decision Theory and Juror Decision-Making*, in *The Trial Process* 353 (Bruce D. Sales ed., 1981); Stuart S. Nagel & Miriam Neef, *Deductive Modeling to Determine an Optimal Jury Size and Fraction Required to Convict*, 1975 Wash. U. L.Q. 933. [\[BACK\]](#)

12. Kenneth R. Hammond, *Human Judgment and Social Policy* 29-30 (1996); Kenneth R. Hammond et

al., *Making Better Use of Scientific Knowledge: Separating Truth from Justice*, 3 Psych. Sci. 80 (1992); Ewart A.C. Thomas & Anthony Hogue, *Apparent Weight of Evidence, Decision Criteria, and Confidence Ratings in Jury Decision Making*, 83 Psych. Rev. 442 (1976). [BACK]

13. Patricia G. Milanich, *Decision Theory and Standards of Proof*, 5 L. & Hum. Behav. 87 (1981). [BACK]

14. Michael L. DeKay, *The Difference Between Blackstone-like Error Ratios and Probabilistic Standards of Proof*, 21 L. & Soc. Inquiry 95 (1996); Ward Edwards, *Influence Diagrams, Bayesian Imperialism, and the Collins Case: an Appeal to Reason*, 13 Cardozo L. Rev. 1025, 1062-65 (1991). [BACK]

15. Allen et al., *supra* note 6, at 195-96; Richard O. Lempert & Stephen A. Salzberg, *A Modern Approach to Evidence: Text, Problems, Transcripts and Cases* 162-63 (2d ed. 1983). Allen et al. also analyzes the civil burden of persuasion in terms of its alleged tendency to "result in about the same number of errors being made for plaintiffs as for defendants." *Id.* at 828. After correctly observing that the preponderance standard has this property only under very restricted "empirical" conditions (*id.* at 829), this text mistakenly concludes that it has translated an "algebraic" proof about maximizing utility into "a geometric representation" about equalizing the numbers of errors. *Id.* at 831. As explained *infra* note 38 and *infra* Part IV(D), the mathematics and the rationale of equalizing expected error rates are quite different from the mathematics and rationale of maximizing expected utility. [BACK]

16. See, e.g., DeKay, *supra* note 13, at 98-99, 127 n.78; D.H. Kaye, *Statistical Significance and the Burden of Persuasion*, 46 L. & Contemp. Probs. 13 (1983) (all discussing *In re Winship*, 397 U.S. 358 (1970), and related cases). [BACK]

17. *Ballew v. Georgia*, 435 U.S. 223, (1978) (opinion of Blackmun, J.). For discussion, see D.H. Kaye, *And Then There Were Twelve: The Supreme Court, Statistical Reasoning, and the Size of the Jury*, 68 Calif. L. Rev. 401 (1980) (criticizing Justice Blackmun's use of statistical decision theory), and authorities cited, *id.* at 1006 n.17. [BACK]

18. See *infra* note 29. [BACK]

19. 509 U.S. 579 (1993). [BACK]

20. See generally David H. Kaye & David Freedman, *Reference Guide on Statistics*, in *Reference Manual on Scientific Evidence* (Federal Judicial Center ed., 2d ed. forthcoming 1998); D.H. Kaye, *Hypothesis Testing in the Courtroom*, in *Contributions to the Theory and Application of Statistics* 331 (Alan E. Gelfand ed., 1987); Kaye, *supra* note 8; Kaye, *supra* note 15. [BACK]

21. But lawmakers have tried. A famous example is a bill that won the unanimous approval of the Indiana House of Representatives in 1897. House Bill No 246 would have given to the state, without royalties, methods to trisect an angle, to square a circle, and to duplicate a cube--three classic feats that are impossible to accomplish in Euclidean geometry. Keith Devlin, *Off Line: Mythical Mathematics*, *The Guardian* (London), July 3, 1997, at 8. The bill languished in the Senate, thanks to the intervention of a mathematician from Purdue University. Associated Press, *Complete History of Indiana Legislature Rolls Off Presses*, *Chicago Tribune*, July 21, 1987, at 3. Writers have thought that the bill defined the value of π to be 3.0, 3.2, and 9.2376. Devlin, *supra* (reporting the biblical value of 3.0); Associated Press, *Today in History*, Jan. 27, 1997 (reporting 3.2); Narendra Jaggi, *A Centenary Celebration of Clear Political Arrogance*, *Pantagraph* (Bloomington Ill.), June 13, 1997, at A12 (citing 45 Proc. Indiana Acad. Sci. 206

(1935) for the view that "the actual legalese of the bill when translated into mathematics gives the value of pi as 9.2376!"). The text of the bill is obscure, but it seems to imply that $\pi = 3.2$. Mark Brader, *Legislating Pi* (visited Sept. 1, 1997) < <http://www.lsa.umich.edu/ling/jlawler/aux/pi.html> > . For additional discussion, see Underwood Dudley, *Mathematical Cranks 192-97* (1992); David Singmaster, *The Legal Values of Pi*, 7 *Math. Intelligencer* 69 (1985). [BACK]

22. Ronald J. Allen, *Rationality, Algorithms, and Juridical Proof: A Preliminary Inquiry*, 1 *Int'l J. Proof & Evid.* 254 (1997). [BACK]

23. Professor Allen is the John Henry Wigmore Professor at the Northwestern University School of Law. Among his many scholarly contributions, he is a coauthor of *Evidence: Text, Cases, and Problems* (2d ed. 1997), and *Constitutional Criminal Procedure: An Examination of the Fourth, Fifth, and Sixth Amendments, and Related Areas* (1995). [BACK]

24. *Id.* at ___. In Professor Allen's terminology, all "formalisms" are "algorithms." *Id.* [BACK]

25. *Id.* at ___. [BACK]

26. *Id.* (footnote omitted). By "base rate," Professor Allen presumably means the proportion of cases in which plaintiffs' allegations that establish liability are true. [BACK]

27. For a cursory statement of this conclusion, see D.H. Kaye, *Statistical Decision Theory and the Burdens of Persuasion: Completeness, Generality, and Utility*, 1 *Int'l J. Evid. & Proof* 313 (1997) (also discussing remarks by Professor Richard Friedman, *Answering the Bayesioskeptical Challenge*, 1 *Int'l J. Proof & Evid.* 276 (1997)). [BACK]

28. *Id.* Professor Allen vehemently denies this. He finds this description of the mathematics "remarkable," "astonishing," and "blind . . . to the deeper implications of the work." Allen, *supra* note 21, at ___. In his opinion, it shows "the algorithm [to be so] bedazzling . . . that the obvious is overlooked." *Id.* at ___ (concluding that "[i]f a policy maker believed these points to be true, adopting Prof. Kaye's preponderance of the evidence standard because of his assertions that 'it is true for all possible base rates' would be a mistake."). Because the derivations are demonstrably true, the exchange resembles the reactions to the solutions to occasional problems in elementary probability theory posed by Marilyn Vos Savant and other popular writers. Whenever Ms. Vos Savant poses a probability puzzler in her *Parade Magazine* column, many of her readers insist--with absolute conviction--on the wrong answer. The most recent instance involved the following problem: "A woman and a man (related) each have two children. At least one of the woman's children is a boy, and the man's older child is a boy. Do the chances that the woman has two boys equal the chances that the man has two boys?" *Id.* at 6; *cf.* Paulos, *supra* note 1, at 64 (posing a similar problem).

The correct answer is "No." It follows trivially from an enumeration of the sample space and the definition of conditional probability. The sample space consists of four possibilities: (boy first, boy second), (boy first, girl second), (girl first, girl second), (girl first, boy second). The man's oldest child is a boy, which eliminates the third and fourth outcomes. Assuming that the children were produced from a random draw of maternal and paternal chromosomes and that abortion or other methods were not used to select for sex, this leaves two equally likely possibilities, and exactly one of those two involves two boys. Hence, the chance that the man has two boys is one-half. The weaker condition that the woman has at least one boy only excludes one possibility--the third. Under the same assumptions, this leaves

three equally likely outcomes, of which exactly one involves two boys. Therefore, the conditional probability for woman the is one-third.

Yet, many readers were outraged that Ms. Vos Savant would say so. One exclaimed: "This is not going to go away until you admit that your are wrong, wrong, wrong!!!" Marilyn Vos Savant, *Ask Marilyn*, Parade Mag., July 27, 1997, at 6 (quoting from a letter from Pearl Meibos). "You are wrong," another complained. "This is borne out by the application of Bayes' rule to the probability structure you imposed, and in the inner refinement functionality as given in the Dempster-Shafer theory of evidential reasoning." *Id.* (quoting from a letter from Dave Ferkinhoff). [\[BACK\]](#)

29. On the meaning of rationality in this context, see, e.g., Helmut Jungerman, *The Two Camps on Rationality*, in *Decision Making Under Uncertainty* 63 (R.W. Scholz ed., 1983); David Kaye, *The Laws of Probability and the Law of the Land*, 47 U. Chi. L. Rev. 34 (1979). For On the relationship between utility (the concept preferred by many decision theorists and economists) and loss (the concept commonly used in theoretical statistics), see, e.g., D.V. Lindley, *Making Decisions* 121-24 (2d ed. 1985); S. James Press, *Bayesian Statistics: Principles, Models, and Applications* 26-27 (1989). Basically, loss is a difference between certain utilities. In the current context, if the utilities of correct decisions are zero, then the two types of quantities are essentially the same. Although Professor Michael DeKay points out that commentators have been too quick to assume that utilities of correct decisions can be set to zero (DeKay, *supra* note 13, at 116-17), I shall assume that the utility of each type of correct decision is the same (which, for ordinal utility functions, is equivalent to setting them to zero). This simplification may be unnecessary, but it makes the exposition slightly easier. [\[BACK\]](#)

30. See, e.g., *Victor v. Nebraska*, 114 S.Ct. 1239 (1994) (holding that a "reasonable doubt" instruction that refers to "moral evidence" and "moral certainty" is consistent with due process). Occasionally, courts speak in more quantitative terms. E.g., *Branion v. Gramly*, 855 F.2d 1256, 1263 n.5 (7th Cir. 1988) ("reasonable doubt means 0.9 or so"); *United States v. Fatico*, 458 F. Supp. 390, 406 (E.D.N.Y. 1978) ("If quantified, the beyond a reasonable doubt standard might be in the range of 95 + % probable."), *aff'd*, 603 F.2d 1053 (2d Cir. 1979). In civil cases, phrases like "preponderance" and "more likely than not" abound, while in some quasi-criminal cases, the proof must be "clear and convincing." See, e.g., *Santosky v. Kramer*, 455 U.S. 745 (1982) (holding that the preponderance standard violates due process when applied to terminate parental rights due to "permanent neglect"); *Addington v. Texas*, 441 U.S. 418, 424 (1979) (holding that due process requires at least proof by clear and convincing evidence rather than a mere preponderance in a "quasi-criminal," involuntary civil commitment); *Agosto-de-Feliciano v. Aponte-Roque*, 889 F.2d 1209, 1220 (1st Cir. 1989); cases cited, *Rivera v. Minnich*, 483 U.S. 582, 584 n.1 (Brennan, J., dissenting). It seems generally agreed that the usual civil preponderance standard means "more probable than not," which means a probability in excess of one-half. See, e.g., *United States v. Fatico*, *supra*, at 403 ("Quantified, the preponderance standard would be 50 + % probable."); *United States v. Shipani*, 289 F. Supp. 43 (E.D.N.Y. 1968), *aff'd*, 414 F.2d 1296 (2d Cir. 1969). The quasi-criminal standards are more difficult to pin down. See, e.g., *Fatico*, *supra*, at 405 ("Quantified, the probabilities might be in the order of above 70% under a clear and convincing evidence burden," while "[i]n terms of percentages, the probabilities for clear, unequivocal and convincing evidence might be in the order of above 80% under this standard."); *United States v. Shonubi*, 895 F. Supp. 460 (E.D.N.Y. 1995), *rev'd*, 103 F.3d 1085 (2d Cir. 1997). [\[BACK\]](#)

31. E.g., Kaplan, *supra* note 3; ; Kaye, *supra* note 8; Lempert, *supra* note 3. [\[BACK\]](#)

32. On the meaning of "subjective" or "personal" probability, see, e.g., Simon French, *Decision Theory:*

An Introduction to the Mathematics of Rationality (1988); Lindley, *supra* note 28; Kaye, *supra* note 28.

[[BACK](#)]

33. This assumes that an all-or-nothing decision must be made: in a criminal case, a defendant found guilty serves a sentence that does not depend on the probability of guilt; in a civil case, a defendant found liable pays damages that do not depend on the probability of liability. Adjusting damages to reflect the probability of liability is discussed in Kaye, *supra* note 6; Levmore, *supra* note 8; Neil Orloff & Jerry Steadinger, *A Framework for Evaluating the Preponderance of the Evidence Standard*, 131 U. Pa. L. Rev. 1159 (1983). [[BACK](#)]

34. Part II contains a proof of some of these statements and an explanation of the statistical terms of art, such as "loss" and "expected value." Arguments over the differential effects of erroneous verdicts for plaintiffs as opposed to defendants lead to judicially sanctioned or mandated departures from the more-probable-than-not standard. *See, e.g.*, *Rivera v. Minnich*, 483 U.S. 582, 584-85 (19__ (Brennan, J., dissenting)). [[BACK](#)]

35. This figure seems to be popular with commentators, largely because Blackstone remarked that "it is better that ten guilty persons escape, than that one innocent suffer." 4 William Blackstone, *Commentaries on the Laws of England* 352 (1769). *See* DeKay, *supra* note 13, at 116. However, other English commentators suggested different trade-offs. *See* Harold J. Berman, *Origins of Historical Jurisprudence: Coke, Selden, Hale*, 103 Yale L.J. 1651, 1706 n.147 (1994) (discussing Hale's ratios); Hammond, *supra* note ?, at 23 (citing Fortesque and Hale); Kaplan, *supra* note 3, at 1077 n.11 (citing Fortesque and Hale). Furthermore, such statements seem to refer to actual error rates rather than relative losses. Actual error rates are influenced by the proportion of guilty persons among those accused of crimes and the conditional probabilities of errors. *See infra* Part IV(D). Therefore, interpreting Blackstone as claiming that the loss for a false conviction is ten times that of the loss for a false acquittal is problematic. *See* DeKay, *supra*; Hammond, *supra*, at 30. Even so, the historic concern articulated as a preference for differential error rates helps motivate the use of a loss function that treats false convictions as more serious than false acquittals. *Id.*; *cf.* Hammond, at 24-25 (discussing this concern as expressed in Judaic writings). [[BACK](#)]

36. *Cf.* M. Granger Morton & Max Henrion, *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis* 25-27 (1990) (listing alternatives). [[BACK](#)]

37. *See* Tribe, *supra* note 3. [[BACK](#)]

38. *See, e.g.*, Richard S. Bell, *Decision Theory and Due Process: A Critique of the Supreme Court's Lawmaking for Burdens of Proof*, 78 J. Crim. L. & Criminology 557 (1987). *But see* DeKay, *supra* note 13 (arguing against this approach). [[BACK](#)]

39. *See* Michael Finkelstein, *Quantitative Methods in Law: Studies in the Application of Mathematical Probability and Statistics to Legal Problems* (1978). *But see* Kaye, *supra* note 5 (questioning this proposal). Judge Richard Posner conflates the $p > \frac{1}{2}$ standard with an error-equalizing standard. Richard A. Posner, *Economic Analysis of Law* 552 (4th ed. 1992) ("This standard . . . implies that of cases decided erroneously, about half will be lost by deserving plaintiffs and about half by deserving defendants."). Professor Allen also slides from the equality of the two types of losses to an equality in the numbers or rates of the two types of errors. Ronald J. Allen, *Burdens of Proof, Uncertainty, and Ambiguity in Modern Legal Discourse*, 17 Harv. J.L. & Pub. Pol'y 627, 641 (1994) ("The conventional understanding of the

burden of persuasion is that . . . its purpose . . . is to allocate errors consistently with our sense of the relevant utilities. In civil cases we want to allocate errors equally over plaintiffs and defendants . . ."). See also Allen et al., *supra* note 6, at 828-31. The latter statement does not follow from the former. See Finkelstein, *supra*; Johnston, *supra* note 9, at 160-61; Kaye, *supra*; *infra* Part IV(D). [BACK]

40. See, e.g., French, *supra* note 31; Leonard J. Savage, *The Foundations of Statistics* (2d ed. 1972); John von Neumann & Oskar Morgenstern, *Theory of Games and Economic Behavior* 26 (3d ed. 1953). [BACK]

41. E.g., French, *supra* note 31; Patrick Maher, *Betting on Theories* 34-83 (1993); Jacob Marschak, *Decision Making: Economic Aspects*, in 1 *International Encyclopedia of Statistics* 116, 124 (William H. Kruskal & Judith M. Tanur eds., 1978). [BACK]

42. E.g., Arthur W. Burks, *Chance, Cause, Reason: An Inquiry into the Nature of Scientific Evidence* 210-12 (1977); Glenn Shafer, *Savage Revisited*, 1 *Stat. Sci.* 463 (1986); Amos Tversky & Daniel Kahneman, *Rational Choice and the Framing of Decisions*, 59 *J. Bus.* S251, S252-54 (1986). [BACK]

43. On the axioms of probability theory, see, e.g., Kaye, *supra* note 28, at 41 n.28; Press, *supra* note 28, at 9-12. Of course, these axioms can be relaxed or reformulated, leading to more general or alternative mathematical systems. See, e.g., Peter C. Fishburn, *The Axioms of Subjective Probability*, 1 *Stat. Sci.* 335 (1986); Patrick Suppes & Mario Zanotti, *Foundations of Probability with Applications: 1974-1995* 3-49 (1996). [BACK]

44. Old ideas continue to be rediscovered, rephrased, or recycled. *Compare* Symposium, 1 *Int'l J. Evid. & Proof* __ (forthcoming 1997), with *Probability and Inference in the Law of Evidence: The Limits and Uses of Bayesianism* (Peter Tillers & Eric D. Green eds., 1988), and Symposium, *Decision and Inference in Litigation*, 13 *Cardozo L. Rev.* 253-1079 (1991). [BACK]

45. See Ball, *supra* note 8, at 817 ("In ordinary actions, the law ignores all the costs and utilities which might be consequences of the judgment except the benefit and loss represented by the sum of money or property awarded or refused."); Posner, *supra* note 9, at 408 ("the standard implicitly equates a dollar lost by someone erroneously adjudged liable to a dollar lost by one erroneously denied compensation."). This construction of a loss function is pursued in Kaye, *supra* note 6. Among other things, that article shows that the expected-loss-minimizing decision rule in the face of uncertainty as to which one of several actors tortiously caused the damage is to impose all the liability on the single tortfeasor who most probably is responsible--even when this probability is less than 1/2. The usefulness of this result is questioned in Steven Shavell, *Economic Analysis of Accident Law* 117 (1987). Although Professor Shavell maintains that in analyzing tort law, we would do better to eschew talk of erroneous verdicts and ask what decision rule best promotes the goal of deterrence (considering the risk of factual or other error in adjudication), other commentators have found the independent focus on the risk of error to be valuable. See Levmore, *supra* note 8, at 696 n.8. [BACK]

46. See, e.g., Johnston, *supra* note 9, at 161 n.36. [BACK]

47. These "process costs" are emphasized in Bruce L. Hay, *Allocating the Burden of Proof*, 72 *Ind. L.J.* 651 (1997). They include the resources spent by each party in producing and presenting evidence. [BACK]

48. If the decisionmaker is risk-neutral, however, utility loss is a linear function of monetary loss, and the same decision rule will emerge as optimal. [BACK]

49. *E.g.*, David Kaye, *Probability Theory Meets Res Ipsa Loquitur*, 77 Mich. L. Rev. 1456, 1468 n. 43 (1977).

[\[BACK\]](#)

50. *See* Posner, *supra* note 38, at 552. [\[BACK\]](#)

51. Regardless of whose utilities are invoked, the loss function need not be linear. For example, in a two-party case one could use a quadratic function that defines the loss to be the square of the dollars left in the wrong pocket instead of the dollars themselves. Orloff & Steadinger, *supra* note 32, at 1165. This would make no difference for present purposes because the loss would be the same (the square) for both parties, and it is this symmetry rather than the precise magnitude of the losses that leads to the more-probable-than-not rule among the class of all-or-nothing rules. *See infra* Part II. [\[BACK\]](#)

52. *But see* Orloff & Steadinger, *supra* note 32 (arguing that the $p > .5$ rule is inferior to a rule that imposes damages weighted by the probability of liability when the loss function is quadratic). The use of a linear loss function is defended in Levmore, *supra* note 8, at 699-704. [\[BACK\]](#)

53. *See* Levmore, *supra* note 8, at 700-01. *But see* DeKay, *supra* note 13 (noting ambiguities in the legal phrases about the convicting the innocent and acquitting the guilty). [\[BACK\]](#)

54. A similar presentation can be found in DeKay, *supra* note 13, at 113. Given the need to explain carefully the logic and implications of Bayesian decision theory before addressing its applicability to legal factfinding, however, such repetition seems advisable. [\[BACK\]](#)

55. I cannot stay home any longer; I must get to the office in the 20 minutes that it takes me to walk there; and I must walk because my car is broken and my bicycle stolen, and so on. Such conditions are required to keep this a simple, binary decision problem. [\[BACK\]](#)

56. An action space is the set of actions from which one will have to choose. *See, e.g.*, Mark J. Schervish, *Theory of Statistics* 144 (1995). [\[BACK\]](#)

57. In the interest of simplicity, this example uses a two-state model of the world. [\[BACK\]](#)

58. That would simplify my task immensely. Knowing the true state of nature (that it is raining), I would have no need to use the expected loss criterion to cope with uncertainty. I would maximize utility by taking the umbrella. [\[BACK\]](#)

59. In these expressions for conditional probabilities, the vertical bar (|) is read as "given" or "conditioned on." For instance, $p(r=0 | \text{data})$ denotes the conditional probability that it will not rain given the observed data. [\[BACK\]](#)

60. Why I should seek to minimize expected loss is discussed in Part IV. [\[BACK\]](#)

61. *See, e.g.*, French, *supra* note 31; Schervish, *supra* note 55. For an application to law, see Kaye, *supra* note 6. [\[BACK\]](#)

62. Ronald J. Allen, *Reasoning and Its Foundation—Some Responses*, 1 Int'l J. Evid. & Proof 343, __ (1997). [\[BACK\]](#)

63. Furthermore, the incorporation of a base rate into the probability judgment gives no support to Professor Allen's characterization of the matter. His claim is that decision rules like (1) are "generally false" because they neglect base rates.

Because of the inherent uncertainty as to the true state of nature, Bayesian decision theory cannot derive valid results about utility maximization by actual "error minimization." Nevertheless, because the actual values of random variables fluctuate about their expected values, there is a statistical connection between minimizing expected losses and minimizing actual losses. See Kaye, *supra* note 5, at 604 n.17. This certainly is not "silliness," but I do not understand Professor Allen to be making this point. He seems more concerned with minimizing actual errors in some direct fashion that attends to "base rates." See *infra* Part IV(D). That too is not mere "silliness," but to appreciate the mathematical results, one must distinguish between the expected values of a random variable (which can be known *ex ante*) and its actual values (which are only accessible *ex post*). [BACK]

64. Allen, *supra* note 61, at ___. [BACK]

65. I return to this point in Part IV. [BACK]

66. Cf. Allen, *supra* note 38, at 629 ("Just as legal amateurs invariably get the law wrong, normally through over-simplification of complex phenomena, my colleagues trained in the law who purport to contribute to debates in other fields invariably fail adequately to appreciate the relevant subtleties."). [BACK]

67. See *supra* note 62. [BACK]

68. The remainder of Professor Allen's paragraph is a variation of the same theme. It builds to a *crescendo* and ends in a *smorzando*:

"But this means that to reduce expected losses, you have to reduce errors, which is exactly the point that Prof. Kaye so severely criticizes me for suggesting. Even more remarkably, after roundly criticizing me for making such a silly point, Prof. Kaye buries away in a footnote exactly the same point: 'For a loss function that gives equal weight to errors favoring plaintiffs and defendants, the expected loss is proportional to the expected number of errors.' 'Directly proportional' would be more accurate, but there is no reason to quibble over words."

Allen, *supra* note 61, at ___. Professor Allen's velitation misses the point. Presenting theorems about *expected values of functions* of the number of errors as if they were theorems about the actual numbers of errors is misguided. By using linear utility functions, one can minimize expected values of both actual numbers and losses simultaneously. See *supra* note 47. This is a convenient simplification, but it does not undermine the fact that the theorems of Bayesian decision theory are general truths inasmuch as they apply to expected values--and not to other quantities. [BACK]

69. Allen, *supra* note 61, at ___. [BACK]

70. His presentation describes an error for a defendant as "offset[ting]" an error for a plaintiff. Perhaps this why he believes that "[i]n civil cases we want to allocate errors equally over plaintiffs and defendants" Allen, *supra* note 38, at 641. See also Allen et al., *supra* note 6, at 829 ("the preponderance standard['s] appointed task [is] equalizing errors among plaintiffs and defendants."). However, the loss

function whose expectation rule (1) minimizes involves no subtraction. *Cf.* Kaye, *supra* note 5 (arguing that a mistake in one case does not compensate for a mistake in another, unrelated case). [BACK]

71. Allen, *supra* note 61, at ___. [BACK]

72. I consider this line of argument in Part IV(D). [BACK]

73. Like Ms. Vos Savant's correspondents who simply could not or would not accept the implications of the definition of conditional probability, Professor Allen does not attend to the necessary technical details. Instead, he persists in his peculiar claims about expected losses, writing: "I can play this game with virtually any burden of persuasion and utility function. For example, I could show how lowering the standard of proof in criminal cases (yes, 'lowering'), no matter what the relative disutility of erroneous verdicts for defendants and the state, could reduce (yes, 'reduce') 'expected losses.' I could also construct a world having the opposite effect." Allen, *supra* note 61, at ___. As we have seen, only the Bayesian decision rule (1) minimizes the expected loss. [BACK]

74. Thus, Professor Allen remarks:

"Prof. Kaye's argument captures nicely one of my basic qualms about the fascination he and others have with the use of algorithms generally to explain or prescribe juridical decision making. In many instances the fascination with algorithms reduces to a belief that juridical decision making can be reduced to procedural methods that are independent of substantive knowledge In this case, he is essentially claiming that we do not need substantive knowledge to know how to set burdens of persuasion in order to optimize our interests; all we need is the 'procedure' of statistical decision theory. He is wrong about that, as he is wrong generally to think that in the juridical context procedural tricks . . . can substitute for substantive knowledge. Juridical decision making requires vast substantive knowledge that, for all the reasons I have given, does not and cannot reduce to procedural methods, at least not the 'procedural method' of Bayes' theorem."

Allen, *supra* note 61, at ___. [BACK]

75. *See, e.g.*, Kaye, *supra* note 26, at ___. [BACK]

"[T]he use of statistical loss functions [is not] a 'formalism,' 'algorithm,' or 'legal theorem' that must do battle against a competing desire for 'judgment in legal affairs.' . . . No 'tension between algorithms and judgment' arises from dissecting or appraising a legal standard that requires judgment to apply. The analysis explains the instructions given to jurors; the jurors must implement them using their best judgment. The mathematics does not diminish the importance of that judgment, but directs attention to how it should be applied." [BACK]

76. A statistician would say that the sample proportion is an unbiased estimator of the population proportion. However, other properties of estimators may be more important in a given application. *E.g.*, David H. Kaye & David Freedman, *Reference Guide on Statistics*, in *Reference Manual on Scientific Evidence* 374 (Federal Judicial Center ed., 1994). [BACK]

77. It is easy enough to find disagreement about the plausibility of certain axioms in the philosophical literature. Many writers find the axioms to be self-evidently true or otherwise compelling, but this reaction is far from universal. *See, e.g.*, authorities cited, *supra* notes 40-41. What is less obvious is what

feature of adjudicative factfinding makes any particular postulate less plausible in law than in other fields. [\[BACK\]](#)

78. On this view, an attribution of probabilities and utilities is correct if it is part of an overall representation of preferences that makes good sense of them and better sense than any competing interpretation. *Id.* at 9. One difficulty with the remarks of jurisprudential skeptics is that they neither offer nor defend any specific competing interpretations. An exception is the proposal of Jonathan Cohen, whose neoteric theory of probability is at least well defined. *See* L. Jonathan Cohen, *The Probable and the Provable* (1977). For discussions of Cohen's system, see, e.g., Kaye, *supra* note 28; *Probability and Inference in the Law of Evidence: The Limits and Uses of Bayesianism* (Peter Tillers & Eric D. Green eds., 1988). [\[BACK\]](#)

79. Maher, *supra* note 40, at 9. A person with the suitable structure of preferences need not consciously assign numerical values to probabilities and utilities; the individual need not even possess the concepts of probability and utility. The claim is that if the preferences meet certain conditions, then they can be reconstructed in this fashion. *Id.* [\[BACK\]](#)

80. Savage, *supra* note 39. Other formulations rarely are mentioned in the legal literature. *See, e.g.,* Kaplan, *supra* note 3, at 1066; Tribe, *supra* note 3. Yet, an earlier, well known version comes from the mathematician Frank Ramsey. Frank P. Ramsey, *The Foundations of mathematics and other logical essays* 58 (1926). Another famous formulation is due to the polymath John Von Neuman and the economist Oscar Morgenstern. Von Neumann & Morgenstern, *supra* note 39. And there are still others. *See* Press, *supra* note 28, at 9-10. [\[BACK\]](#)

81. The formal statement of this independence postulate is more complicated. *See* Maher, *supra* note 40, at 10. [\[BACK\]](#)

82. An additional requirement can be called *normality*. It pertains to preferences with respect to a set of more than two available acts. *See id.* at 21-23. [\[BACK\]](#)

83. This could be Professor Allen's position, for he writes:

"Savage's and Jeffreys' explanation of how probabilities must be formulated in order to employ Bayes' theorem to subjective probabilities does not, as I showed in my paper, map onto trials. Thus, there is no foundation, no axiomatic base, for the application of Bayes' theorem in the trial context, as trials presently are conducted. In any event, that is the point of my paper, which if wrong should be explained to be wrong by the formalists."

Allen, *supra* note ?, at __. However, it is not clear which of Savage's seven postulates Professor Allen believes do not "map onto trials." [\[BACK\]](#)

84. Professor Craig Callen has emphasized this point. *E.g.,* Craig R. Callen, *Cognitive Science and the Sufficiency of "Sufficiency of the Evidence" Tests*, 65 Tul. L. Rev. 1113 (1991); Craig R. Callen, *Notes on a Grand Illusion: Some Limits on the Use of Bayesian Theory in Evidence Law*, 57 Indiana L. Rev. 1 (1982); *cf.* Bell, *supra* note 8, at 564 ("decision theory provides no endorsement of the Court's lawmaking unless factfinders are indeed fully informed and perfectly rational."). *See also* Allen, *supra* note 38, at 642 ("To use Bayes' Theorem requires that one compute the conditional relationships among the pieces of evidence offered to prove some proposition, which results in a combinatorial explosion."); Ronald J. Allen,

Constitutional Adjudication, the Demands of Knowledge, and Epistemological Modesty, 88 Nw. U. L. Rev. 436, 444 (1993) ("To use Bayes's Theorem requires that one compute the conditional relationships among the pieces of evidence offered to prove some proposition, which results in a combinatorial explosion."); Ronald J. Allen, *The Nature of Juridical Proof*, 13 Cardozo L. Rev. 373, 380 (1991) ("humans lack the computational capacity to employ Bayes' Theorem.").

By the same logic, one might argue that judges should not be concerned with the logical consistency of their opinions because they lack the computational capacity to verify consistency by a sequential search through truth tables. *Cf.* Allen, *supra* note 38, at 644 ("How large a belief set could an ideal computer check for consistency in this way? Suppose that each line of the truth table for the conjunction of all these beliefs could be checked in the time a light ray takes to traverse the diameter of a proton . . . and suppose that the computer was permitted to run for twenty billion years, the estimated time from the 'big-bang' dawn of the universe to the present. A belief system containing only 138 logically independent propositions would overwhelm the time resources of this supermachine."). [BACK]

85. Of course, one can question the value of solving highly simplified decision problems (although drastic simplifications of a complex and chaotic phenomena are common enough in the natural sciences). Herbert Simon raised this concern eloquently in many writings well before the same ideas were restated or rediscovered in the legal literature. *E.g.*, Herbert A. Simon, *Reason in Human Affairs* 10-11 (1983):

"Conceptually, the SEU [subjective expected utility] model is a beautiful object deserving a prominent place in Plato's heaven of ideas. But vast difficulties make it impossible to employ it in any literal way in making actual human decisions. . . . SEU theory has never been applied, and never can be applied-- with or without the largest computers--in the real world. Yet, one encounters many purported applications in mathematical economics, statistics, and management science. Examined more closely, these applications retain the formal structure of SEU theory, but substitute for the incredible decision problem postulated in that theory either a highly abstracted utility function and the joint probability distributions of events assumed to be already provided, or a microproblem referring to some tiny, carefully defined and bounded situation carved out of larger real-world reality."

As used to explain or justify the civil burden of persuasion, SEU theory takes the first of these two approaches. A highly abstracted utility function is postulated in light of the goals of adjudication, and probability distributions are assumed to come from jurors, "already provided," as it were, for use with the $p > \frac{1}{2}$ rule. Plainly, this is not a complete and comprehensive theory of factual reasoning. It is a formal structure for handling irreducible uncertainty. The next two subsections argue that the limitations of real jurors and the importance of using epistemologically defensible probabilities in the decision rule do not suggest any better rule for handling the factual uncertainty that remains at the end of trials. [BACK]

86. *See* Friedman, *supra* note 26. The law is different than science in that we rarely continue to gather facts on adjudicated cases. But there are highly publicized exceptions to this generalization. *E.g.*, Gina Kolata, *DNA Tests Are Unlocking Prison Cell Doors*, N.Y. Times, Aug. 5, 1994, at A1; Fox Butterfield, *New DNA Evidence Suggests Sam Sheppard Was Innocent*, N.Y. Times, Feb. 5, 1997, at A7; Desiree F. Hicks, *Blood Sample Sought from Sheppard Suspect*, The Plain Dealer (Cleveland), Feb. 2, 1996, at 1A; Edward Connors et al., *Convicted by Juries, Exonerated by Science: Studies in the Use of DNA Evidence to Establish Innocence After Trial* (1996). In any event, we can have preferences (and hence probability and utility functions) about acts even though we may never learn more about the states of nature that determine the consequences of our acts. [BACK]

87. Because undertaking calculations is an act with its own consequences and costs, it too is subject to the expected utility criterion. I could, for example, apply decision theory to the following set of acts: take the umbrella (without calculating expected loss), leave the umbrella (without calculated expected loss), calculate the expected loss and take the umbrella if and only if that choice minimizes expected loss. See Maher, *supra* note 40, at 6-7. [BACK]

88. Likewise, the impossibility of reducing a judicial opinion to a series of truth tables and verifying seriatim the consistency of all the propositions in the tables hardly implies that judges should feel free to write opinions that are internally inconsistent. Professor Savage offered this advice in contemplating the axioms that he proposed: "So, when certain maxims are presented for your consideration, you must ask yourself whether you try to behave in accordance with them, or to put it differently, how you would react if you noticed yourself violating them." Savage, *supra* note 39, at 7. [BACK]

89. Cf. Richard Lempert, *The New Evidence Scholarship: Analyzing the Process of Proof*, 66 B.U. L. Rev. 439, 453 (1986), reprinted in *Probability and Inference in the Law of Evidence: The Limits and Uses of Bayesianism* 61, 70 (Peter Tillers & Eric D. Green eds., 1988). Subjective probability theory does not necessarily imply that all personal probabilities that are coherent are equally justified. See, e.g., D.H. Kaye, *Do We Need a Calculus of Weight to Understand Proof Beyond a Reasonable Doubt?*, 66 B.U. L. Rev. 657 (1986), reprinted in *Probability and Inference in the Law of Evidence: The Limits and Uses of Bayesianism* 129 (Peter Tillers & Eric D. Green eds., 1988); Maher, *supra* note 40, at 29-33. Since only suitably justified probabilities should be used in computing expected values, there is room to use probabilities other than those given initially (or even finally) by a juror. [BACK]

90. Cf. Ball, *supra* note 8, at 817 ("if we knew that juries had a consistent error on one side, we could decrease the total mistakes by changing the standard to allow for it."). [BACK]

91. We would be departing from the rule that minimizes expected loss with the juror's unmodified subjective probabilities. To the extent that Professor Allen's concern is that this rule might fail to minimize expected loss as seen by a different decisionmaker, he is quite correct. It might fail, and we might want to design the rules of proof accordingly. This point has been stressed by writers who find Bayesian inference valuable in analyzing evidentiary rules. See Lempert, *supra* note 87. [BACK]

92. I would like to think that this what Professor Allen meant when he wrote:

"The reason I can [show that changing the transition probability from the value that already minimizes 'expected losses' further lowers the expected loss] is because the legal system has no interest in a fact finder's subjective expected utility. Rather, its (if I may reify it) concern is the operation of the system as a whole. Thus, it is perfectly understandable that the legal system (that is, those of us who construct it) may disagree with a fact finder's assessment of probabilities, and take action, system-wide, to bring the implications of such assessments in line with our own."

Allen, *supra* note 61, at __; cf. Bell, *supra* note 8, at 568 ("The factors that make trials differ from the decision theorist's ideal are factors that a reasonable lawmaker would consider in establishing rules for standards of proof."). [BACK]

93. Professor Allen thinks that "the algorithms of Prof. Kaye . . . are of limited utility [because] [t]here are an enormous number of incentives operating on litigants in such a way . . . that fact finders' appraisals of probability are skewed in one way or another, and do not result in nice normal curves." *Id.* As shown in

Parts II and III, base rates are irrelevant, except perhaps as they affect judgments of the probability p to be compared to p^* , and normal distributions are not an issue here. [BACK]

94. I am inclined to speculate that jurors often are too prone to convict in criminal cases because they overestimate the probability that the defendant committed the acts alleged, but I would be hard pressed to prove this claim. See Ronald J. Allen, *The Restoration of In re Winship: A Comment on Burdens of Persuasion in Criminal Cases after Patterson v. New York*, 76 Mich. L. Rev. 30 (1977) (complaining that similar intuitions in Barbara D. Underwood, *The Thumb on the Scales of Justice: Burdens of Proof or Persuasion in Criminal Cases*, 86 Yale L.J. 1299 (1977), have not been verified in an empirical study). [BACK]

95. See, e.g., Lempert, *supra* note 3. [BACK]

96. See, e.g., Allen, *supra* note 61, at ___; Allen, *supra* note 92, at 47 n.65 (referring to "the actual effects of choosing one standard of proof over another"); cf. Allen, *supra* note 38, at 641 ("[i]n civil cases we want to allocate errors equally over plaintiffs and defendants "); Allen et al., *supra* note 6, at 828. ("the preponderance of the evidence standard should result in about the same number of errors being made for plaintiffs as for defendants. "). [BACK]

97. Professor Allen implies that if no cases ever had merit, the burden of persuasion should be set at the unattainable value of 1--a jury should never return a plaintiff's verdict. That would be fine, except that some cases do have merit. (Probabilities of 1 are unattainable because the only propositions with probability 1 are tautologies. Propositions about the material world can never be known with absolute certainty. See, e.g., Lindley, *supra* note 28, at 104 (espousing "Cromwell's rule").) If all we had to worry about were situations in which it were known in advance that all cases really should be won by defendants (or all by plaintiffs), then we would have no need for a probabilistic decision rule--or for trials. [BACK]

98. See, e.g., DeKay, *supra* note 13. Professor Allen perspicaciously asserted this conclusion 20 years ago, when he noted that "without knowing the distribution of guilt probability of factually innocent and guilty defendants, we cannot know the actual effects of choosing one standard of proof over another." Allen, *supra* note 92, at 47 n.65. [BACK]

99. The triangular shapes are arbitrary. The same general picture applies to distributions that have other shapes and locations. [BACK]

100. Because the curves do not integrate to one, they are not probability densities. Cf. DeKay, *supra* note 13, at 101 (using a figure with probability densities to indicate the conditional probabilities of false convictions and false acquittals); Allen et al., *supra* note 6, at 829-30 (drawing density curves but stating, inconsistently, that their heights as well as their areas give the "number of trials"). [BACK]

101. Calibration of probability assessments is most easily explained with an example. A weather forecaster gives the probability of rain every day for a year. On 40 days, the forecaster's probability assessments are 0.6, and it actually rains on 24 (60%) of these days. The assessments of 0.6 are well calibrated. For a rigorous and more general definition, see A.P. Dawid, *The Well-Calibrated Bayesian*, 77 J. Am. Stat. Ass'n 605 (1982). For empirical studies, see, e.g., Detlof von Winterfeldt & Ward Edwards, *Decision Analysis and Behavioral Research* 127-31 (1986); Sarah Lichtenstein et al., *Calibration of Probabilities: The State of the Art to 1980*, in *Judgment Under Uncertainty: Heuristics and Biases* 306

(Daniel Kahneman et al. eds., 1982); Elizabeth Loftus & Willem A. Wagenaar, *Lawyers' Predictions of Success*, 28 *Jurimetrics J.* 437 (1988). [\[BACK\]](#)

102. In those $p_i n_i$ cases, the facts really are such that the governing law imposes liability on the defendant. [\[BACK\]](#)

103. *Cf.* Kaye, *supra* note 5, at 605 n.19. [\[BACK\]](#)

104. In this way, one of Professor Allen's concerns applies--at least when one departs from the framework of Bayesian decision theory and tries to justify the more-probable-than-not standard as tending to minimize the incidence of actual errors. As before, the proportion of meritorious claims (the base rate) plays no role in the analysis. [\[BACK\]](#)

105. Some empirical work suggests that for certain tasks, individuals are fairly well calibrated for probabilities near $\frac{1}{2}$, but that more extreme estimates tend to be overstated. *See, e.g.*, Baruch Fischhoff et al., *Knowing with Certainty: The Appropriateness of Extreme Confidence*, 3 *J. Experimental Psych.: Human Perception & Performance* 552 (1977), reprinted in *Judgment and Decision Making* 397, 397 & Figure 24.1 (Hal R. Arkes & Kenneth R. Hammond eds. 1986) ("when people should be right 70% of the time, their 'hit rate' is only 60%; when they are 90% certain, they are only 75% right; and so on."). With regard to the numbers of correct and incorrect decisions, it would not matter that jurors who think that the probability of the facts that generate liability is 90% are wrong about liability 25% of the time. Using the more accurate figure of 75% for the probability of liability would result in the same set of plaintiffs' verdicts--and the same 25% incorrect plaintiffs' verdicts. [\[BACK\]](#)

106. Real jurors, who do not receive feedback on their judgments, probably are not well calibrated, but the extent and direction of their imperfections are unknown. This makes it all but impossible to improve on their judgments by altering the decision rule. [\[BACK\]](#)

107. The efforts of economic analysts to arrive at rules that maximize expected utility by including costs beyond those associated with errors in litigation are a possible exception. *See, e.g.*, Shavell, *supra* note 44; Johnston, *supra* note 9. However, to the extent that the economic models still seek to maximize expected utility, they are applications of Bayesian decision theory. The difference lies in what goes into the utility function; the economists hunt bigger game than the evidence scholars, but they purchase their armaments from the same manufacturer. [\[BACK\]](#)

[DHK](#) [CLASSES](#) [PUBLICATIONS](#) [BIBLIOGRAPHIES](#) [C.V.](#) [LINKS](#)

updated 25 January 2002