1987

Conceptualizing Proof and Calculating Probabilities: A Response to Professor Kaye

Neil B. Cohen
Brooklyn Law School, neil.cohen@brooklaw.edu

Recommended Citation
73 Cornell L. Rev. 78 (1987-1988)

CONCEPTUALIZING PROOF AND CALCULATING PROBABILITIES: A RESPONSE TO PROFESSOR KAYE

Neil B. Cohen†

I

In *Apples and Oranges: Confidence Coefficients and the Burden of Persuasion*,[1] Professor David Kaye has, as usual, presented thought-provoking arguments in a field of increasing interest. His Article is an important contribution to the debate, although, as this Article explores, I disagree with his conclusions.

A. Confidence in Probability

Professor Kaye's Article is primarily a critical analysis of ideas that I initially presented in *Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge*.[2] In that piece I observed that current probabilistic models of proof fail to account satisfactorily for discrepancies between their results and those that the legal system regularly reaches. The proponents of those models (including, but not limited to, Professor Kaye) have often either denied the discrepancies[3] or explained them as the result of policies in the legal system that compete with accuracy in factfinding.[4] I argued, however, that the discrepancies are real and result not from the presence of external policies in the system (though, of course, such policies are present), but rather from a misperception on the part of the models' proponents as to the nature of probabilities determined by legal factfinders.

In particular, I argued that the proponents and their models fail

---

† Professor of Law, Brooklyn Law School. S.B. 1974, Massachusetts Institute of Technology; J.D. 1977, New York University School of Law. The author gratefully acknowledges the support for this Article provided by the Brooklyn Law School Summer Research Fund. In addition, the author thanks Professors D. Michael Risinger, Charles A. Sullivan, Barry Zaretsky, and Michael Zimmer for their assistance.

1 Kaye, *Apples and Oranges: Confidence Coefficients and the Burden of Persuasion*, 73 CORNELL L. REV. 54 (1987).

2 Cohen, *Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge*, 60 N.Y.U. L. REV. 385 (1985).

3 *See* Kaye, *Naked Statistical Evidence* (Book Review), 89 YALE L.J. 601, 610 (1980); Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1348-49 (1971).

4 *See, e.g.*, Kaye, *The Laws of Probability and the Law of the Land*, 47 U. CHI. L. REV. 34, 39-40 (1979); Tribe, *supra* note 3.

to recognize that the legal system can only *estimate* probabilities, rather than determine them with exactitude, and that the implications of this observation require a rethinking of probabilistic formulations of burdens of persuasion. Accordingly, I suggested that we (1) borrow by analogy from the field of statistics the concept of "confidence intervals" or "interval estimates," (2) recognize that conceiving of our estimated probabilities as intervals is more informative than concentrating solely on best guesses, or "point estimates" of the probabilities, and, finally, (3) deem the burden of persuasion satisfied only when the entire confidence interval or interval estimate (rather than just the point estimate) of the probability in question exceeds the relevant threshold.

As an illustration, I presented the hypothetical (and easily quantifiable) case of *High-Tech Supply Co. v. Hacker*, an action for the price of a microchip lost in transit, in which the burden was on the plaintiff-seller to prove by a preponderance of the evidence the merchantability of the chip.[5] The only evidence available to either party was that the chip in question had been selected randomly from a vat containing thousands of chips. Previously, a sample of one hundred chips from the vat had been tested and fifty-one of them were merchantable. Based on this evidence, the best guess, or "point estimate," of the proportion of merchantable chips in the vat and, therefore, of the probability that the chip in question was merchantable, would be 0.51.

I argued, however, that although the point estimate of the relevant probability exceeded 0.5, the plaintiff-seller should lose.[6] The point estimate is just that—an estimate. Even though a majority of the sampled chips were merchantable, it is not necessarily the case that a majority of the chips in the entire vat were merchantable. Yet, it is only if a majority of the chips in the vat were merchantable that it is more likely than not that the particular chip in question was merchantable. Constructing a confidence interval around the point estimate enables us to take into account the nature of the sample in making a judgment as to the entire vat. More specifically, the confidence interval provides additional information enabling us to decide whether to reject the hypothesis that less than a majority of chips were merchantable (and, therefore, the probability that the chip in question was merchantable is equal to or less than 0.5). If that hypothesis is rejected, the alternative—a majority of the chips in question were merchantable (and, therefore, the probability that the chip in question was merchantable exceeds 0.5)—can be accepted and the plaintiff-seller wins the case.

---

5 Cohen, *supra* note 2, at 405-06.

6 *Id.* at 406.

Based on the information from sampling one hundred chips, however, only values below 0.428 could be eliminated as the proportion of merchantable chips in the vat.[7] Therefore, under my theory, the plaintiff-seller has not met its burden of convincing the factfinder by the preponderance of the evidence of the truth of the probandum.

B. Kaye's Critique

It is clear that Professor Kaye disagrees in some way with the theory presented in *Confidence in Probability*, but the nature of his disagreement is less clear. Although early in his critique he describes the underlying idea of *Confidence in Probability* as "incoherent"[8] and an "unholy union,"[9] much of the remainder of his piece proceeds to cast doubt whether Kaye actually challenges my basic thesis. Indeed, Part II of Kaye's article consists largely of a rearticulation of my concept of using interval estimates,[10] which Kaye in *his* articulation apparently finds acceptable, and of a detailed exposition as to determining the appropriate balance between (1) the risk of incorrectly determining that the burden of persuasion has been met (a "false alarm") and (2) the risk of incorrectly determining that the burden of persuasion has not been met (a "miss").[11] This portion of his paper, although disagreeing with the balance of error risks I advocated, again implicitly accepts the underlying theory of *Confidence in Probability*.

Kaye's schizophrenia concerning my theory makes a response difficult. To the extent that he finds the underlying concept acceptable when articulated more consistently with orthodox statistical nomenclature, I am tempted simply to accept his refinements and move on to further applications.[12] I fear, however, that a significant part of his criticism is directed at a misperception of the issue that I addressed in *Confidence in Probability*. Accordingly, I undertake the following more detailed response.

7 This calculation assumes the use of a one-tailed confidence interval at the 95% confidence level. *See id.* Of course, there is nothing sacred about that particular confidence level, and I use it here only as an example. *See id.* at 413-17.

8 Kaye, *supra* note 1, at 57.

9 *Id.*

10 *Id.* at 64-65.

11 *Id.* at 66-73.

12 *See, e.g.*, Cohen, *Taking Factfinding Away from Factfinders: A Probabilistic Perspective* (n.d.) (work in progress).

II

A. Areas of Agreement

It is perhaps easiest to start by observing that there are large areas about which Kaye and I agree. This is no accident; one of my original purposes in writing *Confidence in Probability* was to rebut arguments that the use, as advocated by Kaye and others, of probabilistic methods in the proof process was improper or invalid. I believe that such methods, including probabilistic interpretations of the burden of persuasion, are essentially valid.

Kaye and I both believe that one can speak meaningfully of a probability that a particular event or series of events occurred. We agree, too, that standard probability axioms apply to the determination of that probability. We also agree that the burden of persuasion can be expressed as a probability threshold, i.e., that the burden of persuasion is satisfied when the probability that the probandum occurred exceeds a particular number. Further, we agree that for civil cases governed by the preponderance of the evidence standard the probability threshold is 0.5. We disagree, on the other hand, over subtle but important points in the context of legal proof as to the significance of the amount of information available in making a probability determination, the effect of varying amounts of information on the precision of that determination, and the relevance of the precision.

In recent years, however, the very idea of using probabilistic methods to analyze the proof process has come under attack by commentators who seem to suggest that, although the logic of mathematical methods works in virtually every other human endeavor, it somehow is inapplicable in the special world of law.[13] I believe that these attacks, although valuable stimuli, are fundamentally flawed; they correctly observe certain inadequacies in the traditional application of probabilistic methods of proof, but wrongly conclude from those observations that the methods cannot be applied at all. Demonstrating the error of this argument against standard mathematical logic was one of my goals in *Confidence in Probability.* It is simply not the case that otherwise valid mathematical principles fail to work in the legal environment; rather, the principles will work so long as the problem to which they are applied is properly defined. An analogy to very simple mathematical issues is instructive. Typically, school children first learn the principles of arithmetic in the abstract and then develop the ability to apply these principles to actual situations by the use of so-called "story

[13] *See generally*, L. COHEN, THE PROBABLE AND THE PROVABLE (1977); G. SHAFER, A MATHEMATICAL THEORY OF EVIDENCE (1976).

problems." Often, a child will solve a story problem incorrectly despite having done every arithmetic calculation correctly. The reason for the wrong answer is not that arithmetic does not work in the context of the problem; the reason, of course, is that the child did not properly understand the problem. Had the child understood the problem, he or she would have made different calculations. The conceptualizing of the problem, not the arithmetic, failed.

Similarly, I believe that the inadequacies identified by the critics of probabilistic analysis result not from problems in the mathematics, but in the conceptualization of the legal issue. My point in writing *Confidence in Probability* was to demonstrate that many of the objections raised by those who attack probabilistic interpretations of proof could be alleviated by recognizing and accommodating the imprecise nature of those interpretations.

Thus, in my efforts to defend the foundations of probabilistic analysis of the proof process, I sought to refine its premises. In so doing I phrased my argument largely in somewhat formal, although simple, mathematical terms. An unintended side effect of that choice of writing style was to lead Kaye to devote his attention to my articulation of the mathematical points, rather than the underlying conceptualization of factfinding. Unfortunately this side effect has obscured the areas of substantive agreement and disagreement that Kaye and I have.

I believe that Kaye and I differ less over *how* to calculate than over *what* to calculate. The difference is too important to lose underneath a welter of mathematical hairsplitting. In order to explore my point more completely, and to identify its divergences from Kaye's largely implicit view of the nature of legal proof, let us examine closely the legal system's approach to choice under uncertainty.

B. A Model of Legal Proof

Simply stated, my model of legal proof stresses the importance of taking into account the amount of information utilized in determining the probability estimate. To illustrate the importance of this factor, let us imagine that favorite construct of probability teachers—a gambling machine. This machine, when played, displays a picture of either an apple or an orange; the machine is set to pick randomly the picture to be displayed based on an apples-to-oranges ratio that is programmed into the machine but unknown to the players. The ratio itself is also randomly determined. A player is allowed to watch a number of plays on the machine, and then the player must determine the probability that the next play will result in an apple being displayed. Assume that there is one player on each of three different machines. Player A is allowed to watch three plays, two of which result in apples; she, naturally, picks $2/3$ as the probability of an apple appearing next on her machine. Player B is allowed to observe thirty plays, twenty of which result in apples; she, also, picks $2/3$ as the probability of an apple appearing next on her machine. Finally, Player C is allowed to observe 30,000 plays, 20,000 of which result in apples; not surprisingly, she, too, picks $2/3$ as the appropriate probability for her machine.

Each player has chosen $2/3$ as the probability for her machine, and we would all agree that that choice is the best one in each case given the limited information available. We would not be surprised at all, however, if the secretly programmed probability for machine A were actually quite a bit different from $2/3$, but we would be surprised if the secret ratio for machine C were distant from $2/3$. Indeed, although it would not surprise us if the secret ratio for machine A were actually less than $1/2$, it would surprise us greatly if the secret ratio for machine C were less than $1/2$. Thus, the probability of $2/3$ for machine A connotes something very different than does the probability of $2/3$ for machine C. The model that Kaye has previously advocated, by contrast, treats all three probabilities the same.

For the sake of this argument, let us take the gambling-machine situation one step further. Imagine that after observing the plays on her gambling machine, each player bets against the house that the next play will result in a picture of an apple. Unfortunately, the lights in the room containing the machines fail at the exact moment that those next pictures appear; by the time the lights return, the pictures are absent from the screens. Although no one has seen the crucial pictures, each player is of course confident that her machine showed an apple. Accordingly, when the house declines to pay the players for their alleged winning bet, the players bring suit to recover their claimed winnings.

Assume that the parties to each suit stipulate (1) that the programmed ratio of apples to oranges for the relevant machine was unknown to both parties, (2) that that ratio was randomly selected, and (3) that the results of the plays observed before the bet and the light failure were as detailed above. Assume further that the lack of additional evidence does not itself create a spoliation-like inference against either party. As we know, based on the information available, we would estimate for each machine a probability of $2/3$ that the unobserved picture was an apple.[14]

[14] Of course, as Kaye would correctly point out, the probability is actually a parameter—the secret ratio. We do not know the secret ratio, however; all we can do is estimate it from the information available.

Under the Kaye model, we would simply note, in this case of "justifiably naked statistical evidence,"[15] that $2/3$ is greater than $1/2$, and conclude that each plaintiff has therefore satisfied her burden of persuasion. Thus, despite the sharply differing implications of the varied observations of the three machines, the probabilities in all three cases would be treated as if they were identical. Kaye evidently belongs to the Gertrude Stein school of probability: he believes that a probability is a probability is a probability.

Under my model, on the other hand, we would award a verdict to Player C (20,000 out of 30,000 apples), but not to Player A (two apples in three plays). We might also find for Player B (twenty apples in thirty plays), but with less comfort than we feel in finding for Player C. The reason springs from the observation noted above, that we are more secure in believing that the programmed proportion of apples for Player C's machine is greater than 0.5 than we are that the relevant proportion exceeds 0.5 for the other players' machines.

The difference between Kaye's view and mine arises, I believe, from differences in our concepts of the factfinding process. My concept, apparently unlike Kaye's, incorporates an awareness of two different types of error to which the process is subject. To illustrate my concept, let us imagine a very simple legal dispute. The parties agree on all facts but one. Plaintiff alleges that event X occurred, while defendant alleges that it did not. If X did occur, plaintiff will be entitled to a remedy; if X did not occur, plaintiff will not be so entitled. According to the legal rules governing this dispute, the burden of persuasion is on plaintiff. Plaintiff will prevail if the factfinder determines that plaintiff has demonstrated X by the preponderance of the evidence.

How should the legal system determine whether to give a remedy to plaintiff? We must address this question, which looks to the heart of the factfinding process, before we can appropriately consider competing theories of probabilistic analysis. The question can best be answered after a close analysis of the explicit and implicit steps that constitute legal factfinding.

First, we must recognize that even after all the evidence is presented the factfinder will not (except very rarely) *know* whether X occurred; some doubt will nearly always linger concerning a witness's veracity or accuracy, the trustworthiness of a confession, or something else. At best, the factfinder can seek to determine the probability that X occurred.

---

15 *See* Kaye, *The Limits of the Preponderance of the Evidence Standard: Justifiably Naked Statistical Evidence and Multiple Causation*, 1982 AM. B. FOUND. RES. J. 487.

Second, even though the factfinder cannot be certain whether X occurred, and can only assign a probability to that event, the legal system allows a factfinder to give a remedy to plaintiff in certain cases. Were the legal system to require certainty before awarding a verdict to the party bearing the burden of persuasion, that party could almost never prevail. In the simple hypothetical at hand, the factfinder may give a remedy to plaintiff upon finding it more likely than not that X occurred. In probabilistic terms, we say that the factfinder may find for plaintiff if it finds that the probability of X is greater than 0.5.

Third, the system's willingness to give plaintiff a verdict even though the factfinder does not *know* that X occurred carries the risk of two kinds of error. First, even though the factfinder determines that X more likely than not occurred, X may not, in fact, have occurred. For convenience, let us refer to this first kind of possible error as "fact error." We are willing to accept the risk of fact error because the alternative (i.e., giving a verdict to defendant even when the factfinder believes that the probability that X occurred exceeds 0.5) is worse. If, indeed, X more likely than not occurred, in the long run of similar cases fewer errors will be made if the system awards the verdict to plaintiff than if it awards it to defendant.

A second risk of error is also present: although the factfinder's best single gness based on the evidence presented (E) is that X more likely than not occurred, it is possible if the totality of all available evidence (T)[16] had been presented[17] that the factfinder's best guess would be that it is actually less likely than not that X occurred. Symbolically, while Pr(X)|E is greater than 0.5, Pr(X)|T may be less than 0.5.[18] Recall, for example, that after three plays of the apples-and-oranges machine, player A determined that the next picture displayed was more likely than not to be an apple. Yet, it is quite possible that additional observed plays would have convinced player A that an apple was, in fact, less likely than not to be next. For convenience, let us refer to this second kind of possible error as "probability error."

I believe that, although we are willing to accept the risk of fact

---

16 Save, of course, for that excluded or excludable under the applicable rules of evidence.

17 We assume here that the party's failure to present that information does not support any spoliation-like inference. In other words, to borrow Kaye's terminology, the evidence is "justifiably naked." *See* Kaye, *supra* note 15.

18 In this context, Pr(X)|E means "the probability of X given the evidence presented." Pr(X)|T means "the probability of X given all conceivable relevant evidence." The symbols are Kaye's. *See* Kaye, *Do We Need a Calculus of Weight to Understand Proof beyond a Reasonable Doubt?*, 66 B.U.L. REV. 657, 662-63 n.15 (1986). *See also* Cohen, *The Costs of Acceptability: Blue Buses, Agent Orange, and Aversion to Statistical Evidence*, 66 B.U.L. REV. 563, 568 n.28 (1986).

error, we are less willing to accept the risk of probability error. If probability error occurs, the factfinder will give plaintiff a remedy even though X is less likely than not to have occurred. If it is less likely than not that X occurred, then in the long run of similar cases *more* errors would be made if the verdict were given to plaintiff than if it were given to defendant.

Our aversion to making probability error causes us to be conservative in determining that X more likely than not occurred. We recognize that while there are some cases in which a factfinder's determination, based on available evidence, that the probability of X is greater than 0.5 would probably not be changed by the presentation of additional evidence, there are other cases in which the factfinder's determination, based on the available evidence, that the probability of X exceeds 0.5 might well change with the presentation of additional evidence. Accordingly, before we make a legal finding that X occurred, we seek to be *convinced* that, in fact, X more likely than not occurred.

Convincing the factfinder of such a probabilistic judgment requires more, I believe, than simply noting that the best guess of the probability exceeds 0.5; rather, I believe, the factfinder also takes into account its judgment as to how likely the best guess is to "hold up."[19] Professor Kaye, however, seems to reject this view.[20]

C. Differences from Kaye

Given the foregoing differences in our views of legal factfinding, it becomes apparent that the crux of the difference between Professor Kaye's view of burdens of persuasion and my own is not statistical but conceptual—the difference results from our differing views of legal decisionmaking. The gulf between us is narrower than Kaye seems to believe, but deeper than the differences he raises.

Professor Kaye's article does not address our underlying conceptual differences. Instead, he focuses on statistical methods with which, for the most part, I have no dispute. Again, our disagreement concerns not the mechanics of statistical determination, but rather the nature of what must be determined.

In attacking my theory, Professor Kaye mischaracterizes it, as well as the legal context in which it resides. He mistakenly attrib-

---

19 The risk of nonpersuasion as to probability error falls on the same party bearing the risk of nonpersuasion as to fact error. In our prototypical example, both risks fall on plaintiff; this is appropriate inasmuch as it is plaintiff who is seeking to have the legal system disturb the status quo.

20 *But see* Kaye, *supra* note 18, at 667 n.22 (suggesting possibility of accounting for completeness of evidence by means of second-order probability).

utes to me an intent to "overthrow" accepted probabilistic formulations of the burden of persuasion,[21] and in explicating that point he distorts both my views and the "reigning theory."[22] In particular, Kaye identifies Bayesian decision theory as the reigning theory of probabilistic legal decisionmaking, and calls my own ideas as a "graft"[23] onto that theory which is "incoherent."[24] It is incoherent, according to Kaye, because it "marries the frequentist's confidence coefficient with a subjectivist's posterior probability."[25]

To begin with, Kaye's description of Bayesian inference as the "reigning" view is novel; as recently as 1982, Kaye stated that Bayesian inference is "not used as widely as the classical theories."[26] More important, in making his assertions, Kaye unwarrantedly appropriates exclusively for the Bayesian view important concepts that in fact are common to any system of decisionmaking based on probabilistic information. Most important among these concepts are (1) determination by the factfinder of a probability that the probandum is true, and (2) a threshold number that the probability must exceed in order for the party bearing the burden of persuasion to win his or her case.[27] This appropriation is either trivial—defining all probabilistic legal decisionmaking schemes as Bayesian—or just another salvo in the long-running war of words between many "frequentists" and "Bayesians." In either event, of course, the label Kaye chooses to give any aspect of the decisionmaking process cannot, by itself, have any effect on its validity or usefulness.

Kaye's somewhat apocalyptic description notwithstanding, my thesis, again, is simply that our analysis of factfinding ought to take into account that the factfinder's determination of the probability in question is only an estimate, like any other determination made on the basis of sample data. Recognizing this feature of factfinding in no way alters the status of the estimated probability as a "posterior probability" unless one adopts an unduly narrow definition of that term.

Furthermore, the suggestion that it is wrong to use frequency-based concepts in the context of determining subjective probabilities is curious coming from Professor Kaye. Kaye himself has recog-

---

21 Kaye, *supra* note 1, at 54.

22 *Id.*

23 *Id.* at 57.

24 *Id.*

25 *Id.*

26 Kaye, *The Numbers Game: Statistical Inference in Discrimination Cases* (Book Review), 80 MICH. L. REV. 833, 853 (1982). In all fairness, Kaye did at that time also state that Bayesian inference was becoming "increasingly influential." *Id.* Perhaps the mid-1980s have witnessed the ascendancy to pre-eminence of Bayesian inference.

27 Kaye, *supra* note 1, at 57.

nized the essential equivalence of subjective and frequency-based probabilities by using the two interchangeably,[28] and appropriately so, for, as has long been demonstrated, subjective probabilities obey all the rules of frequency probabilities.[29] Moreover, it is difficult to imagine how subjective probabilities can be understood as anything other than direct analogies to frequency probabilities.[30]

The important difference in reasoning between Professor Kaye and myself, then, is not statistical but conceptual. We differ not on the mathematics of determining probabilities but on the nature of that which is determined. Kaye's almost exclusive attention to the process of calculating probabilities obscures, I believe, the proper focus of analysis: the nature of legal proof and the reasons for deciding, in a world of incomplete knowledge, that a fact is "proven."

III

A. The Contrasting Concepts at Work

Perhaps the difference may best be illustrated by a choice between two gambling analogies.

Imagine the following variation on the apples-and-oranges gamble described earlier. An opaque vat contains a large number[31] of balls, each of which is either black or white. The percentages of each color ball in the vat are not known, however. A number of balls is randomly drawn from the vat and shown to the contestant. The contestant is then given the choice of betting, at even odds, on the color of each ball remaining in the vat.

If twenty balls were drawn from the vat, eighteen of which were white, the contestant's choice is an easy one: bet white. Although the ultimate percentage of white balls in the entire vat may not, in fact, be 90% (eighteen out of twenty), it would be surprising if the

---

28 *See* Kaye, *supra* note 3, at 604.

29 *See* L. SAVAGE, THE FOUNDATIONS OF STATISTICS (1st ed. 1954); *see also* Tribe, *supra* note 3, at 1346-49 (1971).

30 Even though the events to which one assigns subjective probabilities are typically unique (and, therefore, not describable by relative frequencies), at the very least a subjective probability of, for example, one-third conveys the belief that, on the average, one out of three events with that probability will occur.

Kaye also states that the "perceived dissonance" I identify springs from a failure to appreciate the "often overlooked distinction between justified and unjustified naked statistical evidence." Kaye, *supra* note 1, at 56. Kaye, however, fails to appreciate that my conception (as well, one hopes, as all competing conceptions) of burdens of persuasion extends beyond cases in which the evidence is "statistical," be it naked or clothed. The conception applies without regard to the type of evidence. The fact that quantifiable examples are easy to create and understand should not be allowed to obscure this point.

In this light, Kaye's "distinction" amounts to no more than the commonplace (and, by now, unremarkable) observation that sometimes lack of evidence will itself be evidence against the party who could have been expected to produce it.

31 Any number large enough to avoid hypergeometric distortion will do.

percentage were less than 50%. This is important, because so long as the percentage of white balls is greater than 50%, betting on white will, in the long run of the entire vat, result in more wins than losses. Thus, a rational contestant would likely bet on white in this case.

Similarly, if ten balls were drawn from the vat, and all ten were black, then the rational contestant would almost certainly bet black for each remaining ball. Although there may well be a substantial proportion of white balls in the vat, it would be surprising, given the random selection of ten black balls, if the white balls outnumbered the black balls.

Now if 10,000 balls were drawn from the vat, and 5,500 were white, the contestant might pause a bit longer to think before placing his bet. Only 55% of the drawn balls are white; that is a smaller percentage than that relied on in the previous bets. Yet, on reflection, given the large number of drawn balls, it is unlikely that less than a majority of the balls are white. So long as a majority are white, the contestant will win by betting white. Therefore, the contestant will, most likely, bet white. Similarly, a contestant would likely rely on even narrower percentages so long as the sample size is correspondingly larger. If 1,000,000 balls were drawn from the vat, and 510,000 were white, the contestant would still likely bet white. Although only 51% of the drawn balls were white, it would be surprising, given the large number of drawn balls, if less than a majority were white.

Not all cases will be so easy, however. For example, what if three balls are drawn, two of which are black? Or what if five balls are drawn, three of which are white? If the stakes are non-trivial, the contestant is now in a bit of a quandary. In the second drawing, for example, a majority of the balls drawn were white; yet a bet that each of the remaining balls is white carries significant risk. The contestant might still believe that white is the best bet, but it would not surprise him if, in fact, a majority of the balls in the vat were black. A random drawing from a vat with 55% black balls, for example, could easily result in two white balls out of three balls drawn. In effect, the contestant would be saying that although white is a better bet than black, it would not be surprising to lose by betting white.

Most contestants would not choose to risk a substantial portion of their life's savings on this last bet. Indeed, if contestants in the game were given three options—(1) bet white, (2) bet black, or (3) don't bet—it is fairly easy to predict the choice of most contestants in each of the examples described. In the first example, eighteen white balls out of twenty, most contestants would bet white. In the second example, ten black balls out of ten, most contestants would bet black. In the third example (5,500 white balls out of 10,000) and fourth example (510,000 white balls out of 1,000,000), most contestants would bet white. In the fifth example (two black balls out of three) and sixth example (three white balls out of five), however, most contestants would probably opt for choice number three—don't bet.

Under my view of legal decisionmaking, all three choices, including "don't bet," are available to the factfinder. Under Kaye's view, however, the contestant (i.e., factfinder) does not have choice number three available—a bet must be made. This difference is important. In the first four examples, the contestant/factfinder would willingly risk a significant amount of money on the proposition that a particular color of ball is more likely to appear than is another. In the last two examples, by contrast, the contestant/factfinder would not willingly take that risk. To put it another way, in the first four examples the factfinder is convinced that a particular color is more likely than the other color; in the other two examples, the factfinder is not convinced that either color is more likely than the other.

Under my view of legal factfinding, the party who has the burden of persuasion has the burden of convincing the factfinder to make the bet on that party. If the factfinder instead would bet on the other party, *or choose not to bet,* the burden has not been met. In other words, I see three possible responses from the factfinder: (1) bet on white, (2) bet on black, and (3) too close to call. If it is too close to call, the party has not met its burden. Kaye's view, by contrast, forces the factfinder to announce a bet, even if it is one that the factfinder would not voluntarily make and does not offer the factfinder the option of declining the bet because it is too close to call.[32] Kaye apparently believes that the world is very precise.

Let us translate this betting game back into the terms of the prototypical litigation described earlier.[33] "Bet white" is the equivalent of concluding that X is more likely than not. "Bet black" is the equivalent of concluding that X is less likely than not. "Don't bet" is the equivalent of concluding that one cannot safely say either that X is more likely than not or that X is less likely than not. Therefore, if the burden of persuasion is on the party seeking to prove X, that burden is met only if choice 1—"bet white," or "X is more likely than not"—is appropriate. On the other hand, if the burden of persuasion were on the party seeking to prove not-X, the burden

32 Of course, Kaye's theory does, in a narrow sense, allow all three choices, but it allows the third choice—don't bet—only in the rare case of an exact tie. *See* Cohen, *supra* note 2, at 418.

33 *See supra* pp. 84-86.

would be met only if choice 2—"bet black," or "X is less likely than not"—were appropriate.

If choice 3—"don't bet," or "can't safely say whether X is more or less likely than not"—is appropriate, neither party can win if he bears the burden of persuasion. As I explained in *Confidence in Probability*,[34] this situation is the equivalent of equipoise. This expansive concept of equipoise fits more closely the importance given in litigation to determining who bears the burden of persuasion than does the narrower concept apparently favored by Kaye—that there is equipoise only when the probability of X equals exactly 0.5. (Of course, from the viewpoint of the party bearing the burden of persuasion, choices 2 and 3 are identical; that party loses if either of those two choices are made.)

B. The Burden of Persuasion

In order for the foregoing model of legal decisionmaking to be usable, even heuristically, it must provide rules for determining the circumstances in which each choice is appropriate. In choosing these rules, we must remember that any decision rule will necessarily involve risks of error. For our purposes, we must consider two kinds of risks. First, we must consider the risk that choice 1 will be made (and, therefore, plaintiff wins) when the correct choice would have been choice 2 or choice 3 (and, therefore, plaintiff loses). Second, we must consider the risk that choice 2 or 3 will be made (plaintiff loses) when the correct choice would have been choice 1 (plaintiff wins). The first error is commonly known as Type I error (false rejection of the null hypothesis), while the second error is commonly known as Type II error (false failure to reject the null hypothesis). Any decision rule that lessens the risk of Type I error will increase the risk of Type II error, and vice versa.[35] The decision rule chosen should reflect the legal system's judgment as to the optimum balance between the two.

At this point my theory suggests using the concept of confidence intervals or interval estimates in making the appropriate choice, because these intervals relate directly to the risk of Type I error. To be precise, the risk of Type I error equals one minus the confidence level. For example, let us assume that the risk of Type I error in determining that the probability of X exceeds 0.5 (and, therefore, that the plaintiff should win his or her case) is 5% (or

---

34 Cohen, *supra* note 2, at 418-19.

35 As pointed out in *Confidence in Probability*, the decision rule "plaintiff always loses" will reduce the risk of Type I error to zero, but will then obviously entail a very high risk of Type II error. Similarly, the decision rule "plaintiff always wins" will have a zero risk of Type II error, but a high risk of Type I error. *See* Cohen, *supra* note 2, at 414.

0.05).[36] This is the equivalent of saying that if a confidence interval reflecting a 95% confidence level (100% minus 5%) were drawn around the point estimate of the probability, the entire interval would exceed 0.5.

Accordingly, my theory utilizes confidence intervals, or interval estimates, in determining which choice is appropriate. If the entire interval estimate of the probability in question exceeds 0.5, choice 1 is appropriate. If the entire interval is below 0.5, choice 2 is appropriate. If the interval straddles 0.5, choice 3 is appropriate.

Refer back to the game with the apple and orange gambling machine, for example.[37] Player B observed thirty plays, twenty of which resulted in apples; Player C observed 30,000 plays, 20,000 of which resulted in apples. Both players, presumably, would choose ⅔ as the probability. Yet, if one asked Player B if she could comfortably eliminate 0.5 as the actual probability of apples, she would probably answer no. Player C, on the other hand, if asked the same question, would probably answer yes.

Player C's point estimate of ⅔ as the probability is obviously more precise than Player B's point estimate of ⅔. Yet the two point estimates themselves, being identical, do not convey this difference in precision. More exactly, they do not convey the differences between the two estimates as to which other possible values of the probability in question can or cannot be excluded as unlikely to bring about the observed results. In other words, the point estimates do not communicate two players' differing risks of Type I error in excluding possible values in the neighborhood of the point estimate.[38]

A confidence interval or interval estimate, on the other hand, integrates the point estimate with information about the risk of Type I error associated with that point estimate. A confidence interval or interval estimate consists not of just one point, but rather of the point estimate and a range of values surrounding it. The width of the interval is determined by the degree of "confidence" required and by the amount of data available.

Thus although both Player B's and Player C's point estimates of the probability of an apple were ⅔, their interval estimates would differ considerably. At a 95% confidence level, for example, B's interval estimate would be 0.498-0.835, while C's would be 0.661-

---

36 I have picked this threshold because of its familiarity, not hecause of its inherent correctness. I agree with Kaye that there is nothing sacred about the 0.05 signifinance level. *See* Cohen, *supra* note 2, at 412.

37 *See supra* text accompanying note 22.

38 *Cf.* D. BARNES & J. CONLEY, STATISTICAL EVIDENCE IN LITIGATION 125 (1986) ("Every statistic hides something if it condenses a lot of information into a single number.").

0.672. It is, of course, possible that for B's machine, the "true probability" is less than 0.498 or greater than 0.835. Similarly, for C's machine, the true probability could be greater than 0.672 or less than 0.661. The risk that a probability value *outside* these ranges would yield observed results of twenty apples out of thirty (in B's case) or of 20,000 apples out of 30,000 (in C's case) is the risk of Type I error associated with the confidence interval.

As pointed out above, the risk of Type I error associated with excluding values outside the confidence interval is equal to one minus the "confidence level" of the interval; for a 95% confidence interval, for example, the risk of Type I error is 0.05 (1 minus 0.95). Now, recall that the preponderance of the evidence burden of persuasion can be stated probabilistically as the burden of demonstrating that the probability of the probandum exceeds 0.5. Of course, this probability can only be estimated. For the sake of argument, assume that our legal system is willing to accept only a 0.05 risk of Type I error in making the estimate. Under these circumstances, we would award a verdict to the plaintiff only if a 95% confidence interval excluded all values under 0.5 for the probability of the probandum. Only then would the risk of Type I error be within the acceptable range.

As I mentioned above,[39] Kaye is schizophrenic about the use of confidence intervals. Early in his piece,[40] he is quite pejorative; later, however, he rearticulates the theory in a way he finds acceptable.[41] His rearticulation does represent, in many cases, a more appropriate use of certain probabilistic terms of art, and further debate on that matter would serve no good purpose.

C. Optimizing Error Risks

I believe that the confidence interval analogy performs well as a heuristic for the decisionmaking process.[42] To use it, however, a factfinder must select a confidence level to use in constructing the intervals. This necessarily involves balancing the risks of Type I and Type II error in determining whether the probability exceeds 0.5. In *Confidence in Probability*, I evaluated five possible ways that the rela-

39 *See supra* text accompanying notes 9-12.

40 *See* Kaye, *supra* note 1, at 57-62.

41 *Id.* at 64-65.

42 The confidence-interval-based theory is weak, however, in an area that Kaye does not mention. The reason for Kaye's silence, I believe, is that this is the same area in which his theories and all "mathematicist" theories stumble. Most legal disputes involve inherently unquantifiable evidence, consisting of shadings, inconsistent evidence, and the like. Neither Kaye's calculus nor mine can produce a precise number (or numbers) from such evidence that, when compared with a set standard, will tell us who should win the case.

tive risks of these errors could be optimized: (1) minimize Type I error; (2) minimize Type II error; (3) set α, the risk of Type I error, just below 0.5; (4) set β, the risk of Type II error, just below 0.5; and (5) set α equal to β. For reasons described in my article, I chose the fifth alternative—equalizing the risk of Type I and Type II errors.[43]

Professor Kaye believes that my choice is wrong; he would, instead, opt to *minimize* the sum of Type I and Type II errors, a theory he labels δ_{MAP}.[44] His theory is appealing, and ought to have been addressed in *Confidence in Probability*.[45] I do not, however, believe it is necessarily correct or desirable in all cases. In evaluating competing theories as to the balancing of the risk of errors, we must, of course, keep in mind what those errors are. The possible error with which we are concerned here is not fact error[46]—that is, giving the plaintiff a verdict when, in fact, the probandum had not occurred (or vice versa). Rather, we are considering probability error,[47] balancing the relative risk, on the one hand, of giving the plaintiff a verdict when the probability of the probandum is actually less than 0.5, and, on the other hand, of giving the defendant a verdict when the probability of the verdict is actually greater than 0.5.[48]

When my choice, δ_E as Kaye labels it, is close to δ_{MAP}, one might argue along with Kaye that the societal interest in minimizing such factfinder errors outweighs the societal interest in equalizing those errors. When δ_E differs significantly from δ_{MAP}, however, Kaye's point can become less attractive. To suggest the extreme situation, if the method of minimizing the sum of the risks of Type I and Type II errors is to set one risk at zero while letting the other risk rise, it is not clear that δ_{MAP} is a palatable choice.

---

43 Cohen, *supra* note 2, at 417.

44 Kaye, *supra* note 1, at 72.

45 The example Kaye uses to illustrate his theory is inapt, however. In the name of simplicity, Kaye adds so much information to the mythical case of *High-Tech Supply Company v. Hacker* that no serious analyst of the case—Bayesian, frequentist, or otherwise—could reach any conclusion other than that the burden of persuasion is satisfied.

As to Kaye's discussion of lie-detector tests, it shares with my thesis the concept of confidence intervals, but uses them in an entirely different context. The discussion is interesting but inappropriate to analysis of probabilistic definitions of burdens of persuasion. As such, a response is unnecessary in this article.

46 *See supra* text accompanying notes 16-18.

47 *See id.*

48 Kaye misses this distinction in his statement that "[t]he only nonsuperficial analysis of the civil burden of persuasion that I have seen builds on the premise that in civil litigation, a false alarm and a miss are equally serious mistakes." Kaye, *supra* note 1, at 72. Moreover, the statement is an undeservedly harsh assessment of the ideas of other highly competent scholars. *See, e.g.*, M. FINKELSTEIN, QUANTITATIVE METHODS IN LAW: STUDIES IN THE APPLICATION OF MATHEMATICAL PROBABILITY AND STATISTICS TO LEGAL PROBLEMS 69 (1978). *Cf.* Kaplan, *Decision Theory and the Factfinding Process*, 20 STAN. L. REV. 1065, 1072 (1968) ("The assumption [of equal seriousness] is, of course, open to question. Indeed . . . this assumption becomes increasingly dubious.").

For example, assume a class of lawsuits for which, in 40% of the cases, the probability of plaintiff's case[49] exceeds 0.5. Here if the decision rule were "always give plaintiff the verdict," then α, the risk of Type I error, would be 0.6, and β, the risk of Type II error, would equal zero. On the other hand, if the decision rule were "always give defendant the verdict," then α would equal zero, and β would be 0.4. Now assume further that, for this class of lawsuits, the relation between α and β is such that δ_E is satisfied when $\alpha = \beta = 0.24$, and that the point at which the sum of α and β is lowest is when $\alpha = 0$ and $\beta = 0.4$. If we adopt δ_{MAP}, then all the errors will be Type II errors; all verdicts will be for the defendant and all the errors will result from verdicts that should have been for the plaintiff. If on the other hand we adopt δ_E, then plaintiffs and defendants will be equally victimized by errors, and the total number of errors will be only 20% higher (from $\alpha + \beta = 0.4$ to $\alpha + \beta = 0.48$).

It is far from clear that a decision rule that brings about error minimization—but entirely on the backs of the plaintiffs—is preferable here. Accordingly, although Kaye's suggestion of δ_{MAP} is important, more thought is needed before concluding that it is the optimal solution.

IV

I welcome Kaye's critique of *Confidence in Probability*, and find valuable its points concerning both my thesis and its articulation. In retrospect, however, I wish my article had more effectively communicated its focus on the conceptualization rather than the mathematics of legal decisionmaking, for Kaye's critique does not seriously address those underlying concepts.

I look forward to continued debate.

---

49 I.e., Pr(liability) | T. *See supra* note 18.