6-1985

Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge

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

Recommended Citation
60 N. Y. U. L. Rev. 385 (1985)

CONFIDENCE IN PROBABILITY: BURDENS OF PERSUASION IN A WORLD OF IMPERFECT KNOWLEDGE

NEIL B. COHEN*

Commentators and courts have debated with considerable difficulty the correct probabilistic approach to burdens of persuasion. In this Article, Professor Cohen explains why he believes past efforts have failed. He then advances his own probabilistic model of the proof process, relying largely upon the statistical concept of "confidence intervals." Finally, Professor Cohen applies his model to the preponderance of the evidence standard and points out the model's implications for more stringent standards.

INTRODUCTION

In recent years, scholars have hotly debated the role of probabilistic judgments in the legal process. For example, they have employed probability theory to explain and define the concept of relevance[1] and to analyze reapportionment of voting districts.[2] It is the role of probability in legal proof, however, that has generated the most interest and controversy.[3]

Although there are those who strongly advocate the use of probabilistic methods in the proof process, others express doubt about the applicability of traditional probability rules to the determination of facts at issue in a trial.[4] Moreover, the proponents of applying probability theory to legal issues have been unable to articulate clearly the connections among available data, probability calculus, and decisions about liability or guilt.[5] In particular, the current probabilistic models of burdens of persuasion suggest that the preponderance of the evidence standard is met in some important situations in which the legal system traditionally would deny that it is.[6] A close examination reveals that some of these

* Professor of Law, Seton Hall University; Visiting Professor of Law, Brooklyn Law School. S.B., 1974, Massachusetts Institute of Technology; J.D., 1977, New York University. The author would like to acknowledge the contributions made by his colleagues Catherine McCauliff, D. Michael Risinger, Charles A. Sullivan, and Michael Zimmer, all of whom provided helpful criticisms of various drafts of this Article, and by his research assistant, Eileen Mitchell of the Seton Hall University School of Law Class of 1985, whose assistance in all phases of preparation in this Article was exceptional.

[1] See, e.g., R. Eggleston, Evidence, Proof and Probability 79-87 (2d ed. 1983).

[2] See, e.g., M. Finkelstein, Quantitative Methods in Law 105-30 (1978).

[3] See text accompanying notes 10-33 infra.

[4] See text accompanying notes 39-41, 47-52 infra.

[5] See text accompanying notes 67-73 infra.

[6] See text accompanying notes 63-68 infra.

models incorporate a misunderstanding of the nature of the probabilities determined within the legal system.

This Article provides a more theoretically accurate view of these probabilities and proposes an alternative model of their relation to legal proof. Part I summarizes the development of the application of probability theory to legal facts. It sets out the currently accepted probabilistic formulation of the burdens of persuasion and, in particular, of the preponderance of the evidence standard in civil litigation. It then shows how this formulation results in verdicts that are inconsistent with those that actually occur in the legal system and explains why the proffered explanations for the contradictions are inadequate. Part II identifies the incorrect assumptions about forensic probabilities made by the current models and suggests a more theoretically accurate method of analyzing them. It then applies this analysis to construct a more realistic probabilistic formulation of the preponderance of the evidence burden of persuasion. Part III notes implications of the new model for related problems in the proof process.

I

The Current Model

A. *Historical Development*

The process of proving facts in a legal context virtually always involves at least an implicit appeal to the factfinder's intuitive assessment of probabilities. It is not surprising, therefore, that some of the seminal probability theorists—Leibniz in the seventeenth century, Bernoulli in the eighteenth century, and Boole in the nineteenth century[7]—suggested the application of probability calculus in a legal setting. As Professor L. Jonathan Cohen has pointed out, "Leibniz and Bernoulli sought to develop a theory of probability that would have among its principal goals the provision of an adequate analysis for gradations of legal proof."[8]

Although probability theory has flourished for many years as an independent intellectual discipline, scholarly interest in its application within the legal system languished until the latter half of this century. Until relatively recently, the application of probability theory to the law of proof, with some notable exceptions, scarcely had advanced beyond

[7] See G. Leibniz, Allgemeine Untersuchungen Uber Die Analyse Der Begriffe und Wahren Satze (1686); J. Bernoulli, Ars Conjectandi (1741); Boole, On the Application of the Theory of Probabilities to the Question of the Combination of Testimonies on Judgments, 21 Transactions of the Royal Society of Edinburgh 597 (1857); see also L. Cohen, The Probable and the Provable 2, 44, 52 (1977).

[8] L. Cohen, supra note 7, at 2.

Boole's work in the nineteenth century.[9]

Title VII of the Civil Rights Act of 1964,[10] which prohibits employment discrimination based on race, color, religion, sex, or national origin,[11] opened the door to the use of overtly probabilistic and statistical evidence in litigation. Whereas it is easy to win a lawsuit against an employer who explicitly refuses to hire members of one of the protected classes for particular positions, most discrimination is not so blatant. Typically an employer will simply underselect members of the disfavored class without admitting or exhibiting any discriminatory intent. Title VII plaintiffs, therefore, generally must prove their cases through circumstantial evidence.[12] The most common (and often the only) evidence in these cases consists of empirical data about the available labor pool and the work force.[13] Such data, when analyzed through accepted statistical techniques, can provide a reliable basis for drawing inferences about discrimination in the hiring process.[14] Indeed, the Supreme Court has held that empirical evidence from which such statistical inferences can be drawn is sufficient to constitute a prima facie case of illegal discrimination.[15] Therefore, it is important to recognize the effect of Title VII on the development of probability theory in the legal process. Title VII, along with other antidiscrimination statutes such as the Equal Pay Act of 1963[16] and the Voting Rights Act of 1965,[17] provided a powerful incentive for the legal world to develop a greater understanding of the basic principles of probability and statistics. The academic world quickly re-

[9] These exceptions include Ball, The Moment of Truth: Probability Theory and Standards of Proof, 14 Vand. L. Rev. 807 (1961); Cowan, Decision Theory in Law, Science, and Technology, 17 Rutgers L. Rev. 499 (1963); James, Relevancy, Probability and the Law, 29 Calif. L. Rev. 689 (1941); McBaine, Burden of Proof: Degrees of Belief, 32 Calif. L. Rev. 242 (1944); Michael & Adler, The Trial of an Issue of Fact (pts. 1-2), 34 Colum. L. Rev. 1224, 1462 (1934).

[10] 42 U.S.C. §§ 2000e-2000e-17 (1982).

[11] 42 U.S.C. § 2000e-2 (1982).

[12] See, e.g., McDonnell Douglas Corp. v. Green, 411 U.S. 792, 802-07 (1973); Reyes v. Hoffman, 580 F.2d 393, 394-95 (10th Cir. 1978); Calcote v. Texas Educ. Found., 578 F.2d 95, 96-97 (5th Cir. 1978). See generally D. Baldus & J. Cole, Statistical Proof of Discrimination § 1.23, at 44-50 (1980); B. Schlei & P. Grossman, Employment Discrimination Law 1286-1394 (2d ed. 1983).

[13] B. Schlei & P. Grossman, supra note 12, at 156; C. Sullivan, M. Zimmer & R. Richards, Federal Statutory Law of Employment Discrimination 16 (1980).

[14] D. Baldus & J. Cole, supra note 12, § 0.1, at 4-5. See, e.g., International Bhd. of Teamsters v. United States, 431 U.S. 324, 339-40 n.20 (1977).

[15] See Hazelwood School Dist. v. United States, 433 U.S. 299, 307-08 (1977); Griggs v. Duke Power Co., 401 U.S. 424, 426 (1971); see also Lehman v. Trout, 465 U.S. 1056, 1058 (1984) (Stevens, J., dissenting); International Bhd. of Teamsters v. United States, 431 U.S. 324, 339 (1977) (dictum); Movement for Opportunity & Equality v. General Motors Corp., 622 F.2d 1235, 1258 (7th Cir. 1980).

[16] 29 U.S.C. § 206(d) (1982).

[17] 42 U.S.C. §§ 1971, 1973-1973bb(a) (1982 & Supp. I 1983).

sponded not only with a variety of scholarly analyses of statistical and probablistic issues involved in the proof of discrimination,[18] but also with a resurgence of interest in more basic issues regarding the relationship between probability and proof.[19]

The notorious criminal prosecution in the case of *People v. Collins*[20] further stimulated this renewed scholarly interest. In *Collins*, witnesses testified to having seen two individuals—a blond woman with a ponytail and a black man with a beard and mustache—flee together from the scene in a yellow car immediately following a bank robbery.[21] The prosecution attempted to use probability theory to establish that the defendants—a blond woman who had frequently worn a ponytail and a black man who had previously worn a beard and mustache and who drove a yellow car[22]—were the same individuals that the witnesses saw near the scene of the crime. The state called as an expert witness an instructor of mathematics at a state college who testified as to the "product rule," which states, in one form, that the probability that a number of mutually independent events will all occur together is equal to the product of the separate probabilities that each event will occur.[23] The prosecutor then

[18] See, e.g., Baldus & Cole, Quantitative Proof of Intentional Discrimination, 1 Evaluation Q. 53 (1977); Bogen & Falcon, The Use of Racial Statistics in Fair Housing Cases, 34 Md. L. Rev. 59 (1974); Braun, Statistics and the Law: Hypothesis Testing and Its Application to Title VII Cases, 32 Hastings L.J. 59 (1980); Cohn, On the Use of Statistics in Employment Discrimination Cases, 55 Ind. L.J. 493 (1980); Dawson, Are Statisticians Being Fair to Employment Discrimination Plaintiffs?, 21 Jurimetrics J. 1 (1980); Feinberg, The Increasing Sophistication of Statistical Assessments as Evidence in Discrimination Litigation, 77 J. Am. Statistical A. 784 (1982); Finkelstein, The Application of Statistical Decision Theory to the Jury Discrimination Cases, 80 Harv. L. Rev. 338 (1966); Finkelstein, The Judicial Reception of Multiple Regression Studies in Race and Sex Discrimination Cases, 80 Colum. L. Rev. 737 (1980); Grady, Statistics in Employment Discrimination, 30 Lab. L.J. 748 (1979); Hallock, The Numbers Game—The Use and Misuse of Statistics in Civil Rights Litigation, 23 Vill. L. Rev. 5 (1977); Hay, The Use of Statistics to Disprove Employment Discrimination, 29 Lab. L.J. 430 (1978); Kaye, Statistical Evidence of Discrimination, 77 J. Am. Statistical A. 773 (1982); Montlack, Using Statistical Evidence to Enforce the Laws Against Discrimination, 22 Clev. St. L. Rev. 259 (1973); Shoben, Differential Pass-Fail Rates in Employment Testing: Statistical Proof Under Title VII, 91 Harv. L. Rev. 793 (1978); Shoben, In Defense of Disparate Impact Analysis Under Title VII: A Reply to Dr. Cohn, 55 Ind. L.J. 515 (1980); Note, Beyond the Prima Facie Case in Employment Discrimination Law: Statistical Proof and Rebuttal, 89 Harv. L. Rev. 387 (1975); Comment, Judicial Refinement of Statistical Evidence in Title VII Cases, 13 Conn. L. Rev. 515 (1981); Comment, Statistics and Title VII Proof: Prima Facie Case and Rebuttal, 15 Hous. L. Rev. 1030 (1978); Kaye, The Numbers Game: Statistical Inference in Discrimination Cases (Book Review), 80 Mich. L. Rev. 833 (1982).

[19] See, e.g., Kaplan, Decision Theory and the Factfinding Process, 20 Stan. L. Rev. 1065 (1968); Lozowick, Steiner & Miller, Law and Quantitative Multivariate Analysis: An Encounter, 66 Mich. L. Rev. 1641 (1968); Zeisel, Statistics as Legal Evidence, 15 Int'l Encyclopedia Soc. Sci. 246 (D. Sills ed. 1968).

[20] 68 Cal. 2d 319, 438 P.2d 33, 66 Cal. Rptr. 497 (1968).

[21] Id. at 321, 438 P.2d at 34, 66 Cal. Rptr. at 498.

[22] Id. at 322-23, 438 P.2d at 34-35, 66 Cal. Rptr. at 498-99.

[23] Id. at 325, 438 P.2d at 36-37, 66 Cal. Rptr. at 500.

invented hypothetical frequencies with which the characteristics that witnesses had ascribed to the individuals seen leaving the scene of the crime could be expected to occur in the general population,[24] and multiplied these unverified frequencies together to conclude that the probability that a randomly chosen couple would share the distinctive characteristics of the robbers was one in twelve million.[25] Not surprisingly, the defendants, who possessed all of the characteristics, were convicted.[26]

The Supreme Court of California reversed the convictions, holding that the mathematical testimony was inadmissible and that its admission was sufficiently prejudicial to constitute grounds for reversal.[27] The court found the mathematical testimony inadmissible on four grounds. First, the prosecution had supplied no empirical evidence to support the accuracy of the frequencies upon which it based its probability calculation.[28] Second, the product rule in the form used by the prosecution applies only when each factor considered in the equation is independent of all the others—a condition that had not been met in this case.[29] Third, the resulting probability—one in twelve million—assumed that all testimony as to the characteristics of the robbers was accurate—a highly tenuous assumption given the uncertain nature of eyewitness testimony.[30] Fourth, the mathematician had calculated the likelihood that a random couple would possess the combination of characteristics in question—a concept that is quite different from the likelihood that an actual couple possessing those characteristics is innocent.[31]

Although prosecutors in earlier cases had attempted to use the product rule to demonstrate that an accused had committed a crime,[32] the

---

24 Id. at 325, 438 P.2d at 36-37, 66 Cal. Rptr. at 500-01.

25 Id. at 325, 438 P.2d at 37, 66 Cal. Rptr. at 501. The prosecutor thereby implied that there was only one chance in twelve million that defendants were innocent and another couple with the same characteristics committed the crime.

26 Id. at 320-21, 438 P.2d at 33, 66 Cal. Rptr. at 497.

27 Id. at 332-33, 438 P.2d at 41-42, 66 Cal. Rptr. at 505-06.

28 Id. at 327-28, 438 P.2d at 38-39, 66 Cal. Rptr. at 502-03. The prosecutor fabricated the figures and then asked the mathematician to calculate the product. The prosecutor then postulated that the product accurately demonstrated that there was one chance in twelve million that the defendants were innocent and another couple with the same characteristics committed the crime. Indeed, the prosecutor even speculated that his hypothetical figures were conservative and that, as a result, the actual likelihood was much less.

29 Id. at 328-29, 438 P.2d at 39, 66 Cal. Rptr. at 503. For characteristics to be independent of each other, it is necessary that the occurrence of one does not affect the likelihood of any of the others. The court noted that several of the supposedly "independent" variables—such as "Negroes with beards" and "men with mustaches"—obviously represent overlapping categories that are positively correlated with each other.

30 Id. at 330-31, 438 P.2d at 40, 66 Cal. Rptr. at 504.

31 Id. at 331, 438 P.2d at 40-41, 66 Cal. Rptr. at 504-05. The court estimated that the actual probability of guilt was closer to 40%. See id. at 333-35, 438 P.2d at 42-43, 66 Cal. Rptr. at 506-07.

32 See, e.g., State v. Sneed, 76 N.M. 349, 352-53, 414 P.2d 858, 860-61 (1966); People v.

abuse of probability theory in *Collins* inspired a large number of scholars to reflect and comment upon the relationship between probability and proof.[33]

B. *Current Theory*

Today there exists a significant body of literature advocating the application of probability theory to the proof process.[34] Despite conceptual and analytic diversity within the literature, there is general agreement on two fundamental points. First, probabilistic techniques may be used to determine the likelihood of the facts supporting a defendant's guilt or liability.[35] Second, the plaintiff's or prosecutor's burden is satisfied when

Risley, 214 N.Y. 75, 83-85, 108 N.E. 200, 202-03 (1915). Cf. The Howland Will Case, discussed in 4 Am. L. Rev. 625, 648-49 (1870).

33 See, e.g., Broun & Kelly, Playing the Percentages and the Law of Evidence, 1970 U. Ill. L.F. 23; Callen, Notes on a Grand Illusion: Some Limits on the Use of Bayesian Theory in Evidence Law, 57 Ind. L.J. 1 (1982); Charrow & Smith, A Conversation About "A Conversation About *Collins*," 64 Geo. L.J. 669 (1976); Cohen, The Logic of Proof, 1980 Crim. L. Rev. 91; Cohen, Selective Probability and the Paradox of the Gatecrasher, 1981 Ariz. St. L.J. 627; Cullison, Identification by Probabilities and Trial by Arithmetic (A Lesson For Beginners in How To Be Wrong with Greater Precision), 6 Hous. L. Rev. 471 (1969); Cullison, Probability Analysis of Judicial Fact-Finding: A Preliminary Outline of the Subjective Approach, 1 U. Tol. L. Rev. 538 (1969); Fairley & Mosteller, A Conversation About *Collins*, 41 U. Chi. L. Rev. 242 (1974); Feinberg, Teaching the Type I and Type II Errors: The Judicial Process, Am. Statistician, June 1971, at 30; Kaye, The Limits of the Preponderance of the Evidence Standard: Justifiably Naked Statistical Evidence and Multiple Causation, 1982 Am. B. Found. Research J. 487 [hereinafter Kaye, Naked Statistical Evidence]; Kaye, The Paradox of the Gatecrasher and Other Stories, 1979 Ariz. St. L.J. 101 [hereinafter Kaye, The Paradox of the Gatecrashers]; Kaye, Paradoxes, Gedanken Experiments and the Burden of Proof: A Response to Dr. Cohen's Reply, 1981 Ariz. St. L.J. 635; Kornstein, A Bayesian Model of Harmless Error, 5 J. Legal Stud. 121 (1976); Large & Michie, Proving that the Strength of the British Navy Depends on the Number of Old Maids in England: A Comparison of Scientific Proof with Legal Proof, 11 Envtl. L. 555 (1981); Little, Mathematical and Statistical Probability as a Test of Circumstantial Evidence, 19 Case W. Res. L. Rev. 254 (1968); Meier & Zabell, Benjamin Pierce and the Howland Will, 75 J. Am. Statistical A. 497 (1980); Milanich, Decision Theory and Standards of Proof, 5 L. & Hum. Behav. 87 (1981); Orloff & Stedinger, A Framework for Evaluating the Preponderance-of-the-Evidence Standard, 131 U. Pa. L. Rev. 1159 (1983); Underwood, Law and the Crystal Ball: Predicting Behavior with Statistical Inference and Individualized Judgment, 88 Yale L.J. 1408 (1979); Williams, The Mathematics of Proof—I, 1979 Crim. L. Rev. 297 [hereinafter Williams, The Mathematics of Proof]; Williams, A Short Rejoinder, 1980 Crim. L. Rev. 103; Comment, Mathematical Models of Legal Rules: Application, Exploitation and Interpretation, 13 Conn. L. Rev. 33 (1980); Comment, Probability Theory and Constructive Possession of Narcotics: On Finding That Winning Combination, 17 Hous. L. Rev. 541 (1980); Comment, Admissibility of Probability Statistics in a Criminal Trial: *People v. Collins*, 48 Or. L. Rev. 281 (1969); Kaye, Book Review, 89 Yale L.J. 601 (1980) [hereinafter Kaye, Book Review].

34 See, e.g., R. Eggleston, supra note 1; M. Finkelstein, supra note 2; Brook, Inevitable Errors: The Preponderance of the Evidence Standard in Civil Litigation, 18 Tulsa L.J. 79 (1982); Kaplan, supra note 19; Kaye, Naked Statistical Evidence, supra note 33.

35 Much of the work in this area employs Bayes' Theorem to determine probablities. Bayes' Theorem, derived in the eighteenth century by the Reverend Thomas Bayes, allows an initial probability determination to be adjusted to take into account new information. See

that probability exceeds a threshold value.[36]

Although these probabilistic techniques are standard, their application to the determination of legal facts is not straightforward. In traditional probability theory, the probability of a given outcome in some activity is calculated by observing the frequency with which that outcome occurs; the probability value represents the fraction of times that the particular outcome would be expected to occur in a very large or infinite number of repeated trials of the activity.[37] For example, if the number 3 were to turn up on 5000 out of 30,000 throws of a die, the probability of throwing a "3" would be assigned a value of 5000/30,000, or 1/6.

This view of probability—based upon numerous repetitions of the same event—is not directly applicable in determining the likelihood of disputed facts in a legal setting. For example, in the case of a defendant who is being tried for the robbery of a particular bank on a particular date, it would not make sense to discuss the probability that the defendant robbed the bank if by "probability" we really meant "the fraction of all occasions in which this exact evidence was adduced that the defendant actually did rob the bank." The event occurred only once, the fact pattern is unique, and the defendant either robbed the bank or did not rob the bank.

Nonetheless, there is a natural tendency to think in terms of the "probability" that a unique event has occurred. One who did not know that the term "probability" had been reserved, at least formally, for the concept of relative frequency would probably be surprised to learn that the term could not be used to describe the likelihood that a unique event has occurred.

Theory eventually caught up with intuition, and by the 1950's probability estimates of unique events had been incorporated into probability theory. Leonard J. Savage convincingly demonstrated that such estimates, known as "subjective" or "personalist" probabilities, obey all the rules of frequency probabilities, and argued that, as a result, probability calculus could be used to help establish the likelihood of unique events such as those in dispute in a legal setting.[38]

By the mid-1970's, however, an articulate backlash had developed.

Bayes, An Essay Towards Solving a Problem in the Doctrine of Chances (1763). See also T. Wonnacott & R. Wonnacott, Introductory Statistics for Business and Economics § 19-1, at 539-43 (2d ed. 1977) [hereinafter Wonnacott & Wonnacott].

[36] See generally Underwood, The Thumb on the Scales of Justice: Burdens of Persuasion in Criminal Cases, 86 Yale L.J. 1299 (1977). See also United States v. Fatico, 458 F. Supp. 388, 403-11 (E.D.N.Y. 1978), aff'd, 603 F.2d 1053 (2d Cir. 1979), cert. denied, 444 U.S. 1073 (1980).

[37] Wonnacott & Wonnacott, supra note 35, § 3-1.

[38] L. Savage, The Foundations of Statistics (1st ed. 1954).

This movement was led by the British philosopher L. Jonathan Cohen[39] and American law professors Lea Brilmayer and Lewis Kornhauser,[40] who expressed doubt as to the underlying validity of subjective probability calculus, and, therefore, as to the validity of assigning probability values to the unique facts at issue in legal cases.[41] Other scholars, including Glanville Williams,[42] David Kaye,[43] and Richard Eggleston,[44] have rebutted these arguments and have bolstered those supporting the use of subjective probabilities in legal analysis. Today, although the debate as to the validity of the application of subjective probability theory to legal proof is not closed,[45] numerous commentators acknowledge the value of probabilistic interpretation of factual issues involved in a legal dispute.[46] Rather than review and amplify the arguments of that debate, this Article proceeds upon the assumption that probability theory has a legitimate application to problems of legal proof.

Although the views of Cohen, Brilmayer, and Kornhauser have not prevailed, the reservations of another scholar, Laurence Tribe,[47] have had a significant impact. Professor Tribe does not question the theoretical validity of probability calculus in the factfinding process; indeed, he accepts Professor Savage's observation that subjective probabilities obey

---

39 See L. Cohen, supra note 7. Cohen finds that "[p]aradox after paradox emerges from any sufficiently thoroughgoing attempt to represent the logic of legal proof in terms of the calculus of chance." Id. at 2. After examining these anomalies, he concludes that "the mathematicist analysis . . . does not fit the assessment of judicial proof according to existing legal standards and procedures." Id. at 118.

40 Brilmayer & Kornhauser, Quantitative Methods and Legal Decisions (Book Review), 46 U. Chi. L. Rev. 116, 135-48 (1978). The authors criticize the use of Bayesian analysis in legal settings because the theorem does not account for the weight of the evidence, the process of combining issues within the same case, or the concept of the sufficiency of evidence. Id. They conclude that, "[s]ince legal problems are subtle and complex, the unquantifiable variables may well dwarf the quantifiable ones and make numerical modeling futile." Id. at 152-53.

41 See also G. Shafer, A Mathematical Theory of Evidence (1976). Shafer rejects the Bayesian notion that degrees of belief follow the rules of chance. He criticises Bayesian personalists because they "do not seek to analyze the relation between an individual's degrees of belief and his evidence. Nor do they seek to relate the structure of those degrees of belief to the nature of evidence. Instead, they set themselves the task of finding conditions that a set of degrees of belief must obey in order to be internally consistent." Id. at 21.

42 See Williams, The Mathematics of Proof, supra note 33.

43 See Kaye, The Laws of Probability and the Law of the Land, 47 U. Chi. L. Rev. 34 (1979).

44 See R. Eggleston, supra note 1, at 30-49.

45 See, e.g., Nesson, The Evidence or the Event?, 98 Harv. L. Rev. 1357, 1359 (1985) (arguing that "[t]he aim of the factfinding process is not to generate mathematically 'probable' verdicts, but rather to generate acceptable ones [that] will project the underlying legal rule to society and affirm the rule's behaviorial norm").

46 See note 34 supra.

47 Tribe, Trial by Mathematics: Precision and Ritual in the Legal Process, 84 Harv. L. Rev. 1329 (1971) [hereinafter Tribe, Trial by Mathematics]; see also Tribe, A Further Critique of Mathematical Proof, 84 Harv. L. Rev. 1810 (1971).

the principles of objective frequencies.[48] Rather, Tribe argues that "the costs of attempting to integrate mathematics into the factfinding process of a legal trial outweigh the benefits."[49] In particular, Tribe asserts that the explicit use of probabilistic methods often will result in inaccurate and distorted outcomes.[50] Furthermore, he argues that even if more reliable outcomes were produced, such use could undermine values more important to our legal system than those favoring accuracy of factfinding.[51]

Although several commentators have accepted Tribe's thesis,[52] many others continue to advocate the use of probabilistic methods in the legal process.[53] Even if one accepts Tribe's unverified cost-benefit analysis,[54] however, it is critical to note that his criticism is directed at the explicit integration of probabilistic methods into the trial process. He does not directly attack the underlying theoretical validity of applying probability theory to problems of legal proof.

As noted above,[55] the second factor in common among those who advocate the application of probabilistic methods to legal proof is the derivation of a probabilistic formulation of the burden of persuasion, requiring that the defendant be found guilty or liable when the probability of the facts supporting guilt or liability exceeds a particular threshold value.[56] With respect to criminal prosecutions, for which the burden of persuasion is defined as "beyond a reasonable doubt,"[57] courts and commentators do not agree as to the probability of guilt sufficient to convict a defendant[58] or even whether a particular probability value can or should

---

48 Tribe, Trial by Mathematics, supra note 47, at 1346-49.

49 Id. at 1377.

50 Id. at 1358-68.

51 Id. at 1368-77. Tribe identifies three key values in the criminal setting—the presumption of innocence, the notion of innocence until proof of guilt beyond a reasonable doubt, and public faith in the humanity of the legal system—that are threatened by the introduction of probabilistic methods into the factfinding process. Id. These concerns are shared by others. See, e.g., Nesson, supra note 45.

52 See, e.g., Kellman, Anxiety over the TMI Accident: An Essay on NEPA's Limits of Inquiry, 51 Geo. Wash. L. Rev. 219, 234 n.104 (1983); Orloff & Stedinger, supra note 33, at 1162 n.11 (1983); Scobogin, Dangerous Expertise, 133 U. Pa. L. Rev. 97, 124 n.104 (1984); Note, Rules for an Exceptional Class: The Commitment and Release of Persons Acquitted of Violent Crimes by Reason of Insanity, 57 N.Y.U. L. Rev. 281, 313 n.153 (1982); Note, The Inapplicability of Traditional Tort Analysis to Environmental Risks: The Example of Toxic Waste Pollution Victim Compensation, 35 Stan. L. Rev. 575, 584 n.35 (1983).

53 See, e.g., commentaries cited in note 34 supra.

54 This author believes that Tribe's analysis is flawed, especially with respect to civil matters. However, elaboration of this point must await a subsequent article.

55 See text accompanying note 36 supra.

56 See, e.g., Kaplan, supra note 19, at 1071-77.

57 See C. McCormick, Evidence § 344, at 962-64 (3d ed. 1984); 9 J. Wigmore, Evidence § 2497, at 404-15 (J. Chadbourn rev. 1981).

58 See, e.g., United States v. Fatico, 458 F. Supp. 388, 403-11 (E.D.N.Y. 1978), aff'd, 603

be chosen at all.[59]

Proponents of probabilistic decisionmaking generally agree, however, that the preponderance of the evidence standard employed in civil litigation, commonly defined as the burden of demonstrating that the likelihood of the defendant's liability is greater than the likelihood of his innocence,[60] is satisfied by demonstrating that the probability of the existence of the facts supporting liability exceeds 0.5.[61] This standard seems to comport with common sense, and, in fact, a probability of 0.5 has been demonstrated to be the optimal decision point for civil cases, in which "the consequences of an error in one direction are just as serious as the consequences of an error in the other."[62]

In sum, current models applying probabilistic analysis to the proof process share two components: the probability of the event in question is determined by the factfinder, and this probability is then compared with a fixed standard. These models thus share an implicit assumption—that the legal system can and does determine exact probabilities of the facts at issue. This assumption is, however, unwarranted and analytically unfounded. Part II of this Article explores the difficulties associated with this assumption, identifies a more accurate method of describing the ability to assign probability values in the legal system, and formulates a probabilistic definition of the burden of persuasion in civil cases that takes these considerations into account.

II

TOWARD A BETTER THEORY

A. *Theoretical Problems with the Current Model*

In criticizing the application of probability theory to legal proof, Professors Cohen and Tribe identify situations in which the probabilistic

---

F.2d 1053 (2d Cir. 1979), cert. denied, 444 U.S. 1073 (1980) (discussing quantified estimates of burdens of proof in the criminal context); United States v. Schipani, 289 F. Supp. 43 (E.D.N.Y. 1968), aff'd, 414 F.2d 1262 (2d Cir. 1969), cert. denied, 397 U.S. 922 (1970) (discussing the quantification of the preponderance of the evidence standard); Kaplan, supra note 19, at 1073-77 (discussing situations in which different probabilities might suffice to sustain a conviction). See also McCauliff, Burdens of Proof: Degrees of Belief, Quanta of Evidence, or Constitutional Guarantees?, 35 Vand. L. Rev. 1293, 1325 (1982) (author survey showing divergence of opinion among 171 judges concerning the probability of guilt that should suffice to sustain a conviction).

[59] See Tribe, Trial by Mathematics, supra note 47, at 1372-73 (arguing that the criminal justice system should avoid imposing sanctions in the face of a quantitatively measured doubt).

[60] See C. McCormick, supra note 57, § 339, at 956-59.

[61] See, e.g., id.; Kaye, Naked Statistical Evidence, supra note 33, at 493; Kaye, Book Review, supra note 33, at 609; Williams, The Mathematics of Proof, supra note 33, at 297; Winter, The Jury and the Risk of Nonpersuasion, 5 Law & Soc. Rev. 335, 336-39 (1971).

[62] Kaplan, supra note 19, at 1072.

model suggests verdicts inconsistent with current, arguably correct, judicial practice. Professor Cohen presents an example involving a rodeo at which, of 1,000 people in attendance, it is known and uncontested that only 499 paid the admission price—the rest are gatecrashers.[63] Furthermore, there is no objective way to distinguish the gatecrashers from the paying customers. Cohen observes that if the rodeo proprietor brought suit for the admission charge against any particular spectator at the rodeo when the only evidence put to the factfinders was the total attendance and the number in attendance who had paid, the probabilistic model requires that the proprietor recover against each, and presumably against all, of those in attendance because for each case the probability that the defendant is a gatecrasher is equal to 0.501 (which is, of course, greater than 0.5).[64]

Similarly, Tribe posits the case of a tort plaintiff who has been run down by an unidentified blue bus.[65] Tribe suggests that if there is no evidence concerning the identity of the bus owner except that the defendant owns four-fifths of the blue buses in town, the probabilistic model would require a verdict against the defendant because the probability that the defendant is the owner of the bus that hit the plaintiff is 0.8.[66]

Scholarly debate has failed to produce any satisfactory responses to these critical examples. The defenders of probability theory generally do not contend that a verdict for the plaintiff is correct in either case. Rather than questioning the adequacy of their probabilistic decisionmaking model, however, for the most part they have either tried to demonstrate that the probability of liability in the examples did not actually exceed 0.5 or embraced policies other than those of probabilistic decisionmaking to explain the divergence between the verdicts suggested by the model and those that they concede the legal system ought to reach. Tribe, for example, suggests that the very fact that the plaintiff in the bus accident case did not put forth more evidence of ownership of the bus is itself a fact that effectively lowers the probability that one of the defendant's buses hit the plaintiff[67]—an explanation with which Kaye agrees.[68] A similar argument applies to the rodeo case. This argument loses much of its persuasiveness, however, if such additional evidence is

[63] L. Cohen, supra note 7, at 75.

[64] Id.

[65] Tribe, Trial by Mathematics, supra note 47, at 1340-41, 1346-50.

[66] Id. at 1346-48. Tribe argues that this result is inconsistent with the likely real world outcome, in which, without some evidence specifically tying the defendant to the accident, the factfinder probably would fail to find liability. Tribe also notes that the result would not be just because the defendant would be found liable not for four-fifths but for all of the unexplained accidents involving blue buses. Id. at 1349-50.

[67] See id. at 1349.

[68] Kaye, Book Review, supra note 33, at 610.

either unavailable or so difficult to obtain that a party's failure to produce it is not indicative of the likelihood of the facts at issue. Furthermore, Tribe's argument could be applied to the defendant's failure to put forth more evidence as well.

Glanville Williams has suggested that the defendant must win in the rodeo example because, in addition to the probabilistic burden, our sense of justice dictates that the plaintiff should not win if he does not present some evidence specifically tying this particular defendant to the act in question.[69] Although this argument has some surface appeal, a closer examination reveals that its principle is not applicable across a very wide range. For example, if the uncontested facts were that 999 out of the 1,000 rodeo fans were gatecrashers, it is unlikely that our sense of justice would require us to deny judgment to the proprietor because there were no facts other than those overwhelming numbers that tended to show that an individual defendant was one of the 999, and not the one honest customer.[70]

Some theorists have argued that denying judgment to the plaintiff in the bus hypothetical is necessary in order to encourage litigants to find and introduce as much evidence as possible.[71] If plaintiffs such as the hypothetical bus victim could prevail merely by introducing generalized statistical evidence, it is argued, they would have no incentive to develop the facts more fully.[72] Tribe and others find this indicative of the inadequacy of probability theory and of the need to account for other values in the legal system.[73]

These arguments all represent attempts to support the current probabilistic model and, at the same time, to justify deviations from the results that it dictates. None of these attempts, however, withstands close analysis. The arguments must be motivated primarily by a concern with the need to facilitate factually correct trial verdicts. For example, we would not require evidence tying the particular defendant to the act in question unless we believe that this information will more likely result in an accurate decision. Similarly, why should litigants be encouraged to

---

69 Williams, The Mathematics of Proof, supra note 33, at 305.

70 Williams himself acknowledges some doubts as to the analytical consistency of this approach. See id.

71 See, e.g., Kaye, supra note 43, at 39-40; Tribe, Trial by Mathematics, supra note 47, at 1349.

72 See Kaye, supra note 43, at 39 ("the law may exclude mathematical probability calculations precisely to encourage the production of more individualized and revealing evidence"); Tribe, Trial by Mathematics, supra note 47, at 1349 ("[t]o give less [than compelling] force to the plaintiff's . . . omission [of other factual evidence] would eliminate any incentive for plaintiffs to do more than establish the background statistics").

73 See Tribe, Trial by Mathematics, supra note 47, at 1368-77. See also Nesson, supra note 45, at 1359.

find and produce more evidence, if not to enable the factfinder to determine the facts more accurately?

If, however, these arguments reflect a concern that probabilistic methods of proof may lead to inaccurate verdicts, then they are internally inconsistent. For example, in hypotheticals such as those of the rodeo gatecrashers or the blue bus, if it already has already been determined that the probability in question exceeds 0.5, the additional information called for by the arguments to assure an accurate verdict is superfluous under the current probabilistic model of the preponderance standard, which mandates a finding for the plaintiff when he or she demonstrates that the probability of liability exceeds 0.5. Under that model, therefore, there is enough information to find for the plaintiff once it is determined that the probability exceeds 0.5. The assertion that more information is required before finding liability casts doubt on one or both of the two premises of the analysis—the assumption that the evidence already justifies a conclusion that the probability exceeds 0.5, or the view (which is the basis of the currently accepted model) that the plaintiff's burden is satisfied whenever it is demonstrated that the probability of the facts supporting liability exceeds 0.5.

This Article contends that the difficulties encountered by the proponents of the current model in responding to examples such as the gatecrasher and bus cases do not spring from the presence of nonprobabilistic values in the system, such as the need for individualized evidence or the desire to encourage more thorough presentations. Instead, these difficulties result from a misleading probabilistic formulation of the burden of persuasion that is based upon an implicit, and incorrect, assumption concerning the nature of forensically determined probabilities.

B. A More Accurate Theoretical Framework

In the hypothetical situations posited by Cohen and Tribe, the amount of information known about the incident in question is only a small fraction of all of the information that could be known about it.[74]

[74] If one knew all the possible information concerning an event, it would be meaningless to speak of probabilities. As Kaplan implicitly has recognized, if one had perfect information about, say, the placement of a coin on the flipper's thumb, the mass distribution of the coin, air currents, and the force exerted in flipping the coin, one could predict with certainty whether a flipped coin would turn up heads or tails and would not need to resort to the use of probabilities. See Kaplan, supra note 19, at 1066. To state that the probability of heads from a particular flipped coin is 0.5, then, is to concede a certain amount of ignorance.

In the legal context, an analogy can be made to the distinction between circumstantial evidence and "smoking guns." If the factfinder has incontrovertible evidence that a particular event occurred—a smoking gun, so to speak—the probability that the event occurred is 1.0. In the absence of a smoking gun, however, we are left with circumstantial evidence from which only probabilistic judgments can be made.

In other words, the estimated probability of the defendant's liability is based upon only a small amount of data.

This observation illuminates a point that has not been addressed in expositions of probabilistic analysis of proof—that a subjective probability derived by a legal factfinder is more accurately described as an "estimate" based on a small portion, or a "sample," of information rather than as a true value derived from an analysis of all possible information.[75] Thus, in the language of statisticians, a subjective probability determined by a legal factfinder is a point estimate[76] rather than a parameter.[77]

Treating subjective forensic probabilities as point estimates derived from partial information rather than as true values more accurately reflects their nature. As with point estimates, the precision of a subjective probability derived in the legal context is a function of the quantity of information upon which it is based. Treating a subjective legal probability as a sample statistic requires the user to take into account information about the sample size in determining the precision of the estimate and to use that knowledge, along with the point estimate, in making legal decisions.

A nonlegal example will illustrate these principles. Assume that we must determine the likelihood that the next marble drawn from an extremely large vat of black and white marbles will be white, given only the information that out of Y marbles randomly drawn from the vat in the past, X were white. If fifty prior drawings had produced thirty white marbles, our best guess for the probability of the next marble being white would be 0.6. If 100,000 marbles had been drawn, and had yielded 60,000 white marbles, our best estimate of the probability would again be 0.6. Nonetheless, these two probability assessments are quite different; although the estimates are identical, our confidence in their accuracy would differ dramatically due to the difference in the amount of information upon which they were based. In the first case, we would not be surprised if the actual probability of drawing a white marble turned out to be significantly different from 0.6, whereas in the second case a significant deviation from the 0.6 figure would be quite surprising.

This observation is not revolutionary. A professional statistician

75 A few commentators seem to have recognized this point, although they have not incorporated it into their analyses. See, e.g., Kaye, supra note 43, at 52; Lempert, Modeling Relevance, 75 Mich. L. Rev. 1021, 1032 (1977); Tribe, Trial by Mathematics, supra note 47, at 1353-54. Cf. Finkelstein & Fairley, A Bayesian Approach to Identification Evidence, 83 Harv. L. Rev. 489, 493 (1970); Williams, The Mathematics of Proof, supra note 33, at 307.

76 A point estimate is a single estimated figure, or best guess, based on a sampling of the data. See Wonnacott & Wonnacott, supra note 35, § 7-1.

77 A parameter is a constant figure, often with an unknown value, representing the actual status of a population. See id.

would not have described the two probabilities as identical. Rather, he or she would be more likely to say that, based on the information provided in the smaller sample, the probability that the next marble chosen will be white is ninety-five percent certain[78] to be 0.6 plus or minus 0.14 (that is, between 0.46 and 0.74),[79] whereas based on the information provided by the larger sample, the probability is ninety-five percent certain to be 0.6 plus or minus 0.003 (that is, between 0.597 and 0.603).[80] In making these statements, the statistician is describing "interval estimates," or "confidence intervals,"[81] in which it can be said with a specified level of confidence that the true value lies. Notably, the statistician could describe the probability of choosing a white marble in the second case as very likely to be greater than 0.5, but could not do so in the first case.

Returning to the legal context, using these statistical principles we can construct a probabilistic legal decisionmaking model, albeit one that is more complex than that assumed in the current academic writing in the field. This model recognizes that subjective probability assessments by a factfinder are more appropriately analogized to probabilities determined from sample data than to those derived from complete information. The model thus takes into account that the factfinder rarely knows the exact probability of the existence of the event in question and that, rather, the factfinder only can state that he or she is certain, to some degree, that the true probability is within a particular interval, with that interval becoming wider as the degree of certainty required of the factfinder increases. The probabilistic standard for factfinding, therefore, should be described as having a second component. Not only must factfinders determine that their best estimate of the probability in question exceeds the threshold level—0.5 for the preponderance of the evidence standard—based on the evidence presented, but they also must have a certain level of confidence that the true probability, based on all possible evidence, exceeds that threshold.

This view of probability theory as applied to legal proof can be illuminated by returning to nonlegal examples. In the case of the marbles, for example, even though in both examples the estimated probability of choosing a white marble as determined from the available evidence is 0.6,

---

[78] The level of confidence required in any particular situation may be set by the party involved, but it is typically, although arbitrarily, set at 95%. See, e.g., R. Winkler & W. Hays, Statistics: Probability, Inference, and Decision 402-03 (2d ed. 1975); Wonnacott & Wonnacott, supra note 35, § 8-1, at 200.

[79] See method for determining confidence intervals using moderately large samples described in id. § 8.4, at 224.

[80] See method for determining confidence intervals using large samples described in id. § 8.4, at 223.

[81] See id.

only in the second example can one state with a high degree of confidence that the probability of getting a white marble actually is greater than 0.5.[82] If, therefore, the ultimate fact to be proved in a civil trial were whether the next marble chosen will be white, the factfinder could feel that, although the best guess as to the probability of a white marble in both hypothetical cases was 0.6, only in the second case would he or she be confident enough about that guess to find the burden of persuasion satisfied by a preponderance of the evidence.

A number of critical implications concerning the law of proof follow from the application of this concept to the factfinder's determination of liability. First, and most important, it must be recognized that a point estimate, which is an observer's best single estimate of a value such as the probability of liability,[83] conveys much less information than an interval estimate. An interval estimate goes beyond the "best guess" provided by the point estimate. It tells us how precise that guess is by describing a range of values within which one has a particular level of confidence that the true value lies. The point estimate, which does not indicate its precision, gives the user who is not aware of its nature a false sense of exactitude. This ersatz precision can be quite misleading.

A further example illustrates the value of the information conveyed by interval estimates that is not supplied by simple point estimates. Suppose that you are asked to determine whether any of three coins is fair[84] based on the following data: Coin *A* turned up heads 26,000 times in 50,000 flips; coin *B* turned up heads 27 times in 50 flips; and coin *C* turned up heads 51 times in 100 flips. Your best guesses, or point estimates, of the probabilities of heads for coins *A*, *B*, and *C* are 0.52, 0.54, and 0.51, respectively.[85] From the point estimates, it might appear that none of the three coins is fair. However, the point estimates are not necessarily the same as true probabilities. Indeed, the best we can say is that we are confident that the probabilities are somewhat near the point estimates. More particularly, using the data to develop interval estimates at the ninety-five percent confidence level,[86] the probability of heads for coin *A* is 0.52 plus or minus 0.004; for coin *B*, 0.54 plus or minus 0.14; and for coin *C*, 0.51 plus or minus 0.098. The corresponding interval estimate for coin *A*, therefore, is from 0.516 to 0.524; for coin *B* it is from 0.40 to 0.68; and for coin *C* it is from 0.412 to 0.608.

82 At the 95% confidence level, only values below 0.46 can be excluded as the probability of a white marble appearing in the first example. See text accompanying notes 78-79 supra.

83 Wonnacott & Wonnacott, supra note 35, § 20-4, at 573.

84 One would expect a fair coin to turn up heads in 50% of the flips. Therefore, the probability of heads for a fair coin is 0.5.

85 See note 76 and accompanying text supra.

86 See note 78 supra.

When we examine these interval estimates, we see that the only coin that we can state with confidence is not fair is coin *A*. Because the interval estimates for the other two coins straddle 0.5, we cannot confidently exclude 0.5 as a possible true probability of heads for each of those coins.

This example also illustrates an important point about interval estimates. At any particular confidence level, the interval gets smaller as the quantity of data from which the estimate was derived increases. In the example of the three coins, for instance, the confidence interval for coin *A*, for which 50,000 items of information were available, was plus or minus 0.004 from the point estimate. The confidence interval for coin *B*, on the other hand, for which there was information about only fifty flips, was much larger—plus or minus 0.14 from the point estimate.

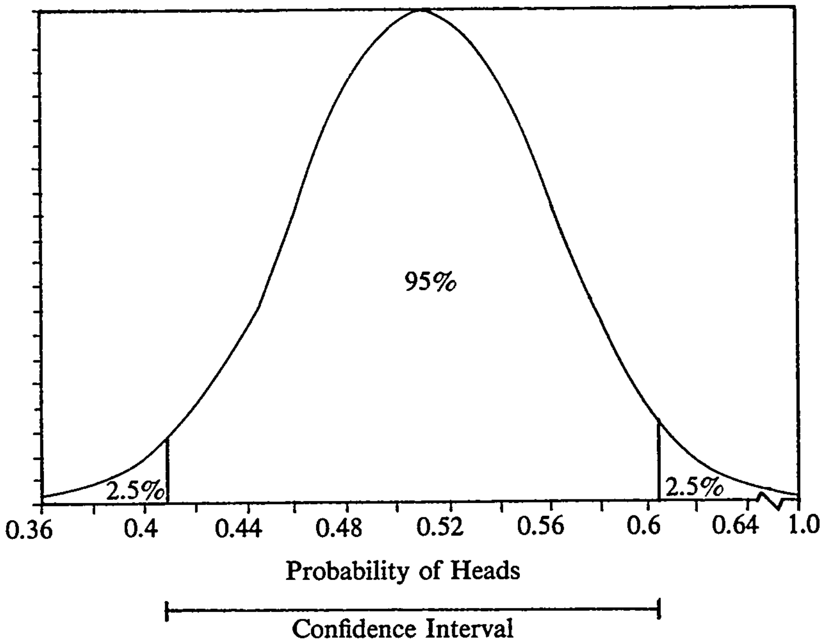
Although the concept of the point estimate as the single best estimate of probability is easily grasped at the intuitive level, quantifying the precision of that estimate by using confidence intervals is conceptually more difficult. Recognizing the existence of a confidence interval concedes a certain amount of ignorance as to the true value of the parameter in question; it is admitting that the true value is not necessarily equal to the point estimate, but is likely to be "in the neighborhood." That "neighborhood" is the confidence interval, and the stipulated level of confidence determines its size. For example, a ninety-five percent confidence interval[87] surrounding a point estimate describes a region in which we believe the true value will fall ninety-five percent of the time.

In the statistician's ideal world of normal distributions and random samples, the concept of confidence intervals can be demonstrated graphically. Figure 1 illustrates how the confidence interval for the probability of heads for coin *C* was determined.

---

87 See id.

FIGURE 1



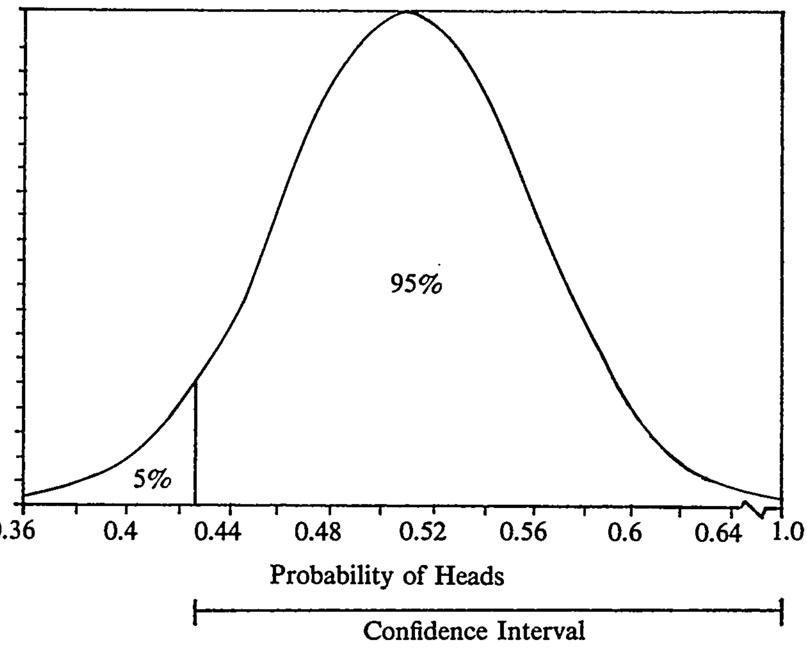
The curve is a "probability curve," indicating the probability of various possible true values for the probability of heads for coin *C*. The total area under the curve, representing the total of all the probabilities of the various possible true values, equals one. The ninety-five percent confidence interval for our estimate of the probability was determined by "cutting off," on each end of the probability curve, a region (or "tail") containing two and one-half percent of the total area under the curve. Because two such tails were cut off, the remaining region contains ninety-five percent of the area under the curve, and therefore represents the region in which the true value will fall ninety-five percent of the time. The confidence interval consists of the probabilities between the tails—in this case, all probabilities between 0.412 and 0.608. Eliminating the highest and lowest extremes excludes from the interval estimates probabilities that are furthest from the point estimate on either side.

In situations in which one is concerned to avoid errors on only one side of the point estimate—for example, when one seeks to establish that a given probability is over a certain threshold but does not care by how much it is over that threshold—calculation of the interval estimate can be further refined. The problem posed by legal proof exemplifies such a situation. The important question in a civil trial is whether the probability of the facts supporting liability exceeds the threshold

amount—how much that probability exceeds the threshold is of no importance. Under the preponderance of the evidence standard, so long as the probability exceeds 0.5, the defendant will be held liable.

In these types of cases, a more appropriate ninety-five percent confidence interval may be constructed by cutting the entire five percent from one tail—here the bottom tail—of the probability curve. Such an interval estimate is commonly referred to as a "one-tailed" estimate,[88] in contrast to the "two-tailed" intervals described earlier. Figure 2 illustrates a one-tailed interval estimate, at the ninety-five percent confidence level, of the probability of heads for coin *C*:[89]

FIGURE 2



A number of important observations concerning probabilities determined by legal factfinders follow from this analysis. First, such determinations are more appropriately treated as sample statistics rather than as true values derived from all possible information. Second, to the extent that factfinders assign a single number to these probability determinations, the determinations are point estimates. As point estimates, they provide only part of the important information that is known about the

[88] Wonnacott & Wonnacott, supra note 35, at § 8-5, at 231.

[89] Notice that, by using the one-tailed test, we have raised the bottom of the interval to 0.428.

likelihood of the issue in question; they represent the factfinder's best guess as to the probability but tell us nothing about the factfinder's confidence in the accuracy of that guess. Knowledge of the confidence interval surrounding the point estimate—at whatever level of confidence is appropriate[90]—gives us more information about the true probability. Finally, as the amount of information used by the factfinder in deriving the point estimate increases, the confidence interval becomes narrower.

These observations lead to the conclusion, upon which the proposed model is based, that we can be confident that the probability of the fact at issue is greater than 0.5 only if the entire interval estimate exceeds that value. Although this observation flows logically from the preceding points, it is a significant change in the customary description of the proof process. Both commentators and courts have tended to assume that factfinders can determine the true values of the probabilities at issue in a case, and typically have described the burden of persuasion in civil cases as the burden of demonstrating that the probability of the existence of the ultimate issue exceeds 0.5.[91] Having only sample information available, however, a factfinder can never determine the true probabilities with absolute precision. The most that a factfinder can do is state that he or she has a particular level of confidence that the probability exceeds 0.5.

C. *Application of the Model to the Proof Process*

Defining the burden of persuasion as the task of demonstrating that the entire interval estimate of the probability of the event in question, rather than merely the point estimate, lies in the region greater than 0.5 complicates the determination of whether the burden has been satisfied. The plaintiff still must convince the factfinder that his or her best guess of the true probability is greater than 0.5. In addition, however, the factfinder must now be convinced that this determination would likely recur if all possible information were presented at trial, or if another fairly selected subset of information were taken.

The coin flipping examples presented earlier[92] demonstrate this point. Our best guess of the probability of heads for coin *B*, which turned up heads twenty-seven times in fifty tosses, is 0.54, yet we are not

---

[90] See note 78 supra.

[91] See, e.g., United States v. Fatico, 458 F. Supp. 388, 403-04 (E.D.N.Y. 1978), 603 F.2d 1053 (2d Cir. 1979), cert. denied, 444 U.S. 1073 (1980); United States v. Schipani, 289 F. Supp. 43, 55-56 (E.D.N.Y. 1968), aff'd, 414 F.2d 1262 (2d Cir. 1969), cert. denied, 397 U.S. 922 (1970). See also text accompanying notes 60-61 supra. But see M. Finkelstein, supra note 2, at 59-78 (arguing that the equalization of errors between parties may require a higher probability threshold—i.e., more than 0.5—than does the minimization of errors).

[92] See text accompanying notes 84-87 supra.

confident that if we tossed this coin an infinite number of times the proportion of heads would in fact exceed 0.5. With respect to coin *A*, on the other hand, although our point estimate of the probability of heads was only 0.52—closer to 0.5 than that of coin *B*—we are confident that if we flipped coin *A* an infinite number of times the proportion of heads would exceed 0.5. This confidence is the result of the large number of flips made.[93] Thus, if a litigant bore the burden of persuading a factfinder that each coin was unfair, that burden would be met only for coin *A*. Although the point estimate of the probability of heads for coin *B* is higher, a factfinder could not be confident that the true probability is indeed greater than 0.5.

A more realistic example illustrates this point in a legal context. Suppose that Hacker, an electronic hobbyist, contracted to purchase from High-Tech Supply Company, a mail order firm specializing in electronic supplies, one Z99 silicon chip for two hundred dollars, F.O.B. High-Tech's warehouse.[94] Within a reasonable time, High-Tech tendered to a carrier a package addressed to Hacker containing a Z99 chip. Unfortunately, through no fault of any party, the package was lost in transit. Nevertheless, High-Tech, thankful that its attorney had recommended that all mail-order contracts of the company be F.O.B. High-Tech, sent a bill to Hacker for the chip. Not surprisingly, Hacker refused to pay for the chip that she had never received. High-Tech, insensitive to consumer goodwill, filed suit against Hacker for two hundred dollars.[95]

In her answer, Hacker admitted the existence of the contract but denied knowledge as to whether the chip was merchantable.[96] She further stipulated that the chip had, in fact, been tendered to the carrier. Therefore, the only factual issue to be resolved at trial was whether the chip was merchantable.

At trial, High-Tech presented testimony concerning its acquisition of the Z99 chips. High-Tech did not manufacture Z99 chips; its entire supply consisted of one thousand chips that were purchased "as is" at an auction of the inventory of a bankrupt electronics manufacturer. The auctioneer admitted that the manufacturer had experienced a serious quality control problem and that the lot of one thousand chips had not

[93] See text accompanying note 87 supra.

[94] See U.C.C. § 2-319 (1978).

[95] High-Tech was relying on U.C.C. § 2-509(1)(a) (1978), which transfers the risk of loss in shipment contracts to the buyer upon delivery of the goods to the carrier, and on id. § 2-709(1)(a), which enables sellers to recover the price of conforming goods lost or damaged within a reasonable time after risk of loss has passed to the buyer.

[96] Because High-Tech is considered a merchant under the Code, see id. § 2-104(1), the chip had to have been merchantable to conform to the contract. See id. § 2-314.

been tested, but he made the chips available for inspection by potential purchasers.

High-Tech's representative at the auction tested one hundred randomly selected chips and found that fifty-one of the one hundred chips were in working order and forty-nine were defective. Despite the large percentage of defective chips in the sample, High-Tech's representative bid on and purchased the lot. High-Tech did not inspect them further. The chips were advertised unsuccessfully in High-Tech's catalogue for several months; indeed, Hacker was the first to order one of the chips. The chip tendered to the carrier for delivery to Hacker was chosen randomly from the lot, and the following day High-Tech sold the remaining chips to a scrap dealer who subsequently destroyed them.

Hacker presented no evidence at trial and stipulated to the accuracy of High-Tech's testimony. Therefore, the sole question for the factfinder is whether the undisputed facts demonstrate by a preponderance of the evidence that the chip intended for Hacker was merchantable. In other words, the factfinder must assess the probability that the chip in question was merchantable.

Based on the evidence, our best guess of the probability of merchantability is 0.51 because fifty-one of the one hundred tested chips were satisfactory. That guess, however, is only a point estimate. An interval estimate provides additional information. A one-tailed interval estimate of the probability that the chip was merchantable demonstrates that, at the ninety-five percent confidence level, we can only say that the probability exceeds 0.428.[97]

In sum, although the factfinder's best estimate of the probability that the chip was merchantable is greater than 0.5, he or she cannot state with sufficient confidence—assuming that we adopt the ninety-five percent confidence level—that the probability actually exceeds that number. Under the current probabilistic model, the point estimate of 0.51 would dictate that High-Tech should win. If, on the other hand, the burden of persuasion is defined so as to require the plaintiff to demonstrate that the entire interval estimate of the probability exceeds 0.5, High-Tech would lose.

Describing the burden of persuasion in terms of confidence intervals or interval estimates is more consistent with the way factfinders and courts intuitively operate[98] than is the traditional method, which assumes knowledge of true values. Furthermore, defining the burden of persuasion in this way enables us to explain, in probabilistic terms, cases such as

[97] The one-tailed interval estimate has its lowest point at $\bar{X} - \bar{Z}_{05}\sigma\bar{X}$, and its highest point at 1.0. See Wonnacott & Wonnacott, supra note 35, § 8-2, at 210, and § 8-5, at 231.

[98] See note 143 infra.

the gatecrashers and bus company, which are otherwise difficult to reconcile with prior concepts of probabilistic proof. In both cases it can easily be argued that, although the point estimate of the probability of liability is greater than 0.5, the confidence intervals are so wide that they straddle 0.5. The result is that we cannot say with sufficient confidence that the true probability of liability exceeds 0.5.

This point is not obvious at first. In both cases, we are given full population data rather than just samplings. We are given information about all the spectators at the rodeo and all of the buses in town. Thus, at least superficially, it appears that the true probability that any random spectator was a gatecrasher equals 0.501 and that the true probability that any random blue bus was operated by the Blue Bus Company equals 0.8. However, those probabilities are not at issue. The issue in Cohen's example is not "What is the probability that a randomly selected spectator at the rodeo was a gatecrasher?" Rather, the question at issue is "What is the probability that this particular defendant was a gatecrasher?" Absent from this determination is a great deal of information beyond the overall proportion of gatecrashers in the rodeo crowd. For example, does the defendant have a witness who can testify that he or she saw the defendant purchase a ticket? What is the defendant's reputation for honesty? Such information would be relevant to any estimate of the probability that the defendant was a gatecrasher. Indeed, according to Bayes' Theorem,[99] answers to these questions could be combined with the information we already have about the number of gatecrashers to derive a modified, or "posterior," probability of the defendant's liability. That these questions are unanswered creates our uncertainty. Tribe and Kaye came close to articulating this point when they suggested that the plaintiff's failure to produce more evidence amounted to evidence against the plaintiff.[100] However, their observation is unsatisfactory because the same could be said about the defendants, who also failed to produce evidence. It is more accurate to say that the failure of both parties to produce more evidence gives rise to a large degree of uncertainty. We realize that there are many more relevant but unknown facts about the incident in question, but we do not know which side they favor—that is, we do not know if they would increase or decrease our probability estimate. What we do know is that additional information would result in a more precise estimate.

Once again, our analogy to coin flips is illustrative. If we flip a coin about which we know nothing two or three times with heads resulting each time, our best guess of the probability of heads for that coin will

[99] See note 35 supra.

[100] See text accompanying notes 67-68 supra.

exceed 0.5. Yet no one would confidently state that the probability of heads for that coin actually does exceed 0.5. Why? Because we are intuitively aware that even a fair coin occasionally will produce the same result the first two or three times it is tossed. Before being able to state anything about the coin with confidence, we would want to know more information. The missing information, which could either confirm that the coin is unfair or suggest that it is indeed fair, is the source of the uncertainty.

Similarly, in the case of the rodeo gatecrasher,[101] we know that missing information about the defendant and the incident could alter substantially our probability estimate. Returning to statistical terminology, although our point estimate of the probability that a randomly selected defendant was a gatecrasher is 0.501, our interval estimate of that probability straddles 0.5. Accordingly, we are unable to state with confidence that the probability that this particular defendant is a gatecrasher exceeds 0.5. Thus, with the burden of persuasion defined as the task of demonstrating that the interval estimate of the probability of the defendant's liability lies entirely in the region above 0.5, the plaintiff rodeo proprietor will lose its case against the spectator, just as Professor Cohen believes it should.

Professor Tribe's bus case[102] can be explained similarly. Again, although our point estimate based on the available evidence is 0.8, there is so much missing evidence that is relevant to the probability of the bus company's liability that here, too, the confidence interval could straddle 0.5.[103] Therefore, the quantum of proof necessary to prevail according to the model presented in this Article could again yield a verdict against the plaintiff, consistent both with Tribe's suggestion and with the likely result in court.

Note that the proposed model allows us to use insights about confidence intervals even when the evidence does not allow the size of the interval to be quantified, as is typically the case. The model therefore serves a heuristic role by illuminating how factfinders apply probabilty theory in actuality and by demonstrating the effect of limited amounts of information upon the factfinding process. With respect to the rodeo gatecrasher and blue bus cases, for example, even though we cannot determine the exact width of the confidence intervals in question, we can conclude that they are wide enough to straddle 0.5. Thus, this model

101 See text accompanying notes 63-64 supra.

102 See text accompanying notes 65-68 supra.

103 Examples of the missing information might include the safety records of the company and its competitors, information as to whether the company operated buses on the road or street where plaintiff was injured, and testimony from the drivers on duty that night.

explains why a factfinder would reach a result different from the one suggested by the traditional model, which uses only the point estimate.

D. *Ascertaining the Proper Level of Confidence*

1. *The Relationship Between Level of Confidence and the Risk of Errors*

The essential difference between the model of probabilistic proof proposed in this Article and previous models is the distinction this model recognizes between point estimates and interval estimates. Both models define the burden of persuasion, in cases governed by the preponderance of the evidence standard, as the requirement of convincing the factfinder that the probability of the defendant's liability exceeds 0.5. The new model, however, recognizes that a factfinder cannot determine the probability exactly but instead can only estimate it with a certain degree of confidence. Accordingly, this model reformulates the burden of persuasion as that of demonstrating that it can be stated with a particular level of confidence that the probability in question is greater than 0.5. The level of confidence chosen will determine the width of the interval surrounding the point estimate of probability.[104] The prior model, on the other hand, operates as though the factfinder can determine the true probability of defendant's liability. In other words, the prior model assumes that the factfinder is one hundred percent confident of his or her probability estimate, with a confidence interval of zero width surrounding the estimate.

Although the prior model's assumption makes that model easier to understand, it also makes the model highly unrealistic in the real world of incomplete knowledge, because the width of the interval estimate necessarily increases with the level of confidence demanded.[105] A confidence level of one hundred percent would result in an interval estimate so wide that even the most trivial probabilities would not be excluded.[106] Conversely, an interval estimate coterminous with the point estimate would

[104] See text accompanying notes 74-81 supra. See also Wonnacott & Wonnacott, supra note 35, § 8-1(a), at 200

[105] See authorities cited in note 104 supra. See also Wonnacott & Wonnacott, supra note 35, § 9-4(a), at 260.

[106] An interval estimate representing 100% confidence will foreclose only impossible results. For example, if a coin which has been flipped a number of times with both heads and tails resulting, the interval estimate, at the 100% confidence level, for the probability of heads will exclude only 1.0 (because some tails have occurred) and 0.0 (because some heads have occurred). However, even if the coin had been flipped hundreds or even thousands of times with heads being the only result, the 100% confidence interval would not exclude the possibility, however remote, that tails might occur in the future.

have an infinitesimally small confidence level.[107]

The difficult, but critical, question presented by the new model is ascertaining the level of confidence that the legal system should require in constructing the interval estimates used in the preponderance of the evidence test. Because any level of confidence involves a risk of error, choosing the appropriate confidence level requires consideration of both the possibility of errors and the costs we place on them.

Two possible errors may result from a decision based on sample information: a party who should not be found liable may be found liable (a "false inculpation"), and a party who should be found liable may be found not liable (a "false exculpation").[108] Statisticians have labeled false inculpation as "Type I error" and false exculpation as "Type II error."[109]

When we express a particular level of confidence in a probabilistic determination, we are making a statement about the risk of Type I (false inculpation) error associated with that determination. The relationship between the two concepts is direct: the risk of Type I error, usually represented by the Greek letter α,[110] is equal to one minus the confidence level. Thus,

$$\alpha = 1 - \text{confidence level}$$

and

$$\text{confidence level} = 1 - \alpha.$$ [111]

Choosing a confidence level of ninety-five percent is thus the equivalent of accepting a five percent risk of Type I error. If the preponderance of the evidence standard is defined as the burden of demonstrating that the interval estimate of the probability of defendant's liability at a ninety-five

107 Because the confidence level is equal to the fraction of the area under the probability curve that is contained in the confidence interval, it can be mathematically described as:

$$C = \int_a^b p(x)dx$$

Clearly, $\lim_{a \to b} C = 0$

108 In other words, a defendant whose probability of liability is equal to or less than 0.5 may nonetheless be found liable, and vice versa. See Wonnacott & Wonnacott, supra note 35, § 9-4(a), at 259.

109 Id. The risk of Type I error, usually represented by the symbol α, is defined more generally in hypothesis testing, see note 117 infra, as the error of rejecting the null hypothesis when it is true. The risk of Type II error, usually represented by the symbol β, is the risk of accepting the null hypothesis when it is false. Id.; H. Blalock, Social Statistics 113-14 (2d ed. 1972).

In the legal context, the null hypothesis is that the defendant is not guilty or not liable. If, at the chosen level of confidence, the factfinder determines that the defendant is liable (rejects the null hypothesis) when in fact the defendant should not be held liable, a Type I error has been made. Conversely, if the factfinder determines that a defendant is not liable (accepts the null hypothesis) when he should in fact be held liable, a Type II error has occurred.

110 Wonnacott & Wonnacott, supra note 35, § 9-3(c), at 255-56.

111 Id.

percent confidence level lies entirely in the region greater than 0.5, then that standard incorporates a five percent risk that a defendant for whom the actual probability of liability is less than 0.5 will lose nonetheless. In other words, at the ninety-five percent confidence level, five percent of all defendants who would win if their true probability of liability were known will incorrectly lose.

The relationship between the risk of Type I error and the risk of Type II error, and hence between the confidence level and Type II error, is not easily ascertainable.[112] Although the two risks are inversely related in that increasing one decreases the other,[113] they are not simple complements—that is, they do not add up to one. Indeed, in situations in which the parameter being estimated can take on a full range of values having an unknown distribution, such as the probability of a defendant's liability or guilt, the exact risk of Type II error cannot be determined directly.[114]

It is possible to approximate the risk of Type II error, however, by making assumptions about the distribution of the true probabilities and constructing models based on these assumptions.[115] Using such methods, one statistician, for example, has created a model of an employment discrimination case in which setting the risk of Type I error at five percent (and thus the confidence level at ninety-five percent) resulted in a risk of Type II error equal to approximately fifty percent.[116]

---

112 See Kaye, Statistical Significance and the Burden of Persuasion, 46 Law & Contemp. Probs. 13, 22 (Autumn 1983) (citing M. DeGroot, Probability and Statistics 373-82 (1975)).

113 It is easy to see, for any given quantity of data, that a rule of decision that decreases the likelihood of Type I (false inculpation) errors will increase the likelihood of Type II (false exculpation) errors, and vice versa. At the extremes, for example, a rule dictating that all defendants be found liable would have a risk of Type II error equal to 0 but would also have a very high risk of Type I error. Conversely, a rule dictating that defendants never be found liable would have a risk of Type I error equal to 0 but a high risk of Type II error. If all defendants were found liable, there could be no situations of mistaken exculpation. Similarly, if no defendant were ever found liable, false inculpation could never occur. Wonnacott & Wonnacott, supra note 35, § 9-4(b), at 260.

114 Wonnacott & Wonnacott, supra note 35, § 9-5, at 263. Because β varies with the true probability, the risk of Type II error in a test cannot be quantified unless we know something about the distribution of true probabilities. Id. at 264-67. We can, however, assume a set of true values of the probability in the liability range and then for each such value calculate the corresponding value of β, the risk that the confidence level chosen will result in an incorrect exculpation. By subtracting these results from 1.0, one can calculate for each possible probability in the liability range the probability that the defendant correctly will be found liable. The set of these probabilities of correct inculpations usually is referred to as the "power function" of a test, and a graph of the function often is called a "power curve."

115 See H. Blalock, supra note 109, at 244-49; Dawson, Probabilities and Prejudice in Establishing Statistical Inferences, 13 Jurimetrics J. 191, 201-09 (1973) (approximating β for analogous chi-square sample testing method).

116 See Dawson, Investigation of Fact—The Role of the Statistician, 11 Forum 896, 907-08 (1976).

It is conventional in both the physical and the social sciences to use the ninety-five percent confidence level in hypothesis testing[117] and in constructing confidence intervals.[118] This convention reflects nothing more than an arbitrary balancing of the disutilities, or "regrets,"[119] of Type I and Type II errors.[120] It represents a value judgment within the context of those types of research as to the relative costs of incorrectly proclaiming a result on one hand and incorrectly deeming a result not to have been demonstrated on the other.[121] Researchers in these areas have chosen to accept a relatively high risk of Type II error (that of failing to reject the null hypothesis, and, therefore, of not accepting a proposition that is in fact true) in order to minimize the risk of Type I error (that of incorrectly rejecting the null hypothesis, and, therefore, of accepting a false proposition). Although this conservative balancing of risks may be appropriate for deciding when to accept a scientific hypothesis, it is not necessarily appropriate within the legal context. Before accepting this or any other convention for the legal standard of proof, its balancing of the risks of Type I and Type II error—and, therefore, the relative disutilities or regrets implicitly assigned to them—must be examined to determine if that balancing reflects appropriate social judgments in the context of civil litigation.

Unfortunately, the legal system generally has applied, without critical analysis, the conventional but arbitrary test for scientific hypotheses to factual propositions based on empirical data in civil lawsuits. For ex-

117 Readers familiar with statistical methods may have observed that the description of the burden of persuasion in terms of confidence intervals is related to the statistical technique known as "hypothesis testing." See Wonnacott & Wonnacott, supra note 35, § 9-6, at 271-72. Indeed, the two techniques are equivalent. Id. § 8.2, at 207-12. In hypothesis testing, the complement of the proposition sought to be established is tentatively assumed, and given the label "null hypothesis" or "H_o." Wonnacott & Wonnacott, supra note 35, § 9-1, at 244-45. See also note 109 supra. Then, the sample data are compared with a probability distribution of H_o. In classical hypothesis testing, a confidence level—again, typically 95%—is subtracted from 1.0 to define an "acceptance level." If the likelihood of achieving the sample data given the null hypothesis is less than the acceptance level, the null hypothesis is deemed "disproved" at that level and the "alternative hypothesis"—that is, the hypothesis that the null hypothesis is false—is accepted. Id. § 9-3, at 252-56. Just as with interval estimates, there are "one-tailed" and "two-tailed" hypothesis tests. Id. § 9-6, at 269-71. In the context of legal proof subject to the preponderance of the evidence standard, the null hypothesis would be that the likelihood of the facts supporting liability is less than or equal to 0.5. If that hypothesis is deemed disproved, the alternative hypothesis—that the probability exceeds 0.5—is accepted.

118 In a two-tailed test, the resulting interval will extend 1.96 standard deviations on either side of the point estimate. See Wonnacott & Wonnacott, supra note 35, § 9-6(b). This distance is often rounded to 2.0 standard deviations in constructing the interval estimate. See, e.g., id. § 8-1, at 200.

119 See, e.g., Kaplan, supra note 19, at 1078-79.

120 Wonnacott & Wonnacott, supra note 35, § 9-4 (discussing Type I and Type II errors in the context of a criminal trial).

121 See id.; Dawson, supra note 18, at 2.

ample, in *Castaneda v. Partida*,[122] a case involving alleged discrimination against Mexican-Americans in jury selection, the Supreme Court stated, without explanation, that "As a general rule for such large samples, if the difference between the expected value and the observed number is greater than two or three standard deviations, then the hypothesis that the jury drawing was random would be suspect to a social scientist."[123] Similarly, the Supreme Court has adopted, without articulating a justification, the methodology of the *Castaneda* decision in employment discrimination cases under Title VII.[124]

Merely to borrow a standard from the scientific world without examining the values implicit in such a standard is a mistake. This point applies not only to proof of discrimination,[125] but to broader issues in legal proof as well.[126] The Court's lack of analysis of this issue is an abdication of the responsibility to determine an appropriate allocation of risks.

2. *Choosing the Appropriate Level of Confidence in Civil Litigation*

We must recognize that in choosing a confidence level of ninety-five percent, we are selecting a level which will result in defendants being found liable in five percent of the situations in which the true probability of their liability does not exceed 0.5. It also implies that defendants will be found not liable in perhaps fifty percent of the cases in which the true probability of their liability is greater than 0.5.[127] In other words, the fraction of defendants who are exonerated but who would be found liable if the factfinder knew the true probabilities is many times higher than the fraction of those who are held liable but who would be found innocent if the factfinder knew the true probabilities. Such a standard assumes that the social disutility of wrongful inculpation is many times greater than

---

122 430 U.S. 482 (1977).

123 Id. at 497 n.17. Two standard deviations translates into a confidence level of just over 95%. See note 118 supra.

124 See Hazelwood School Dist. v. United States, 433 U.S. 299, 311 n.17 (1977). See also, e.g., Davis v. City of Dallas, 483 F. Supp. 54, 58 (N.D. Tex. 1979) (finding a "statistically significant disparity"—6.31 standard deviations—between the expected number of black police officers and the actual number hired).

125 E.g., Kaye, supra note 112, at 14 n.7.

126 Discrimination cases involve the testing of hypotheses that are different from those addressed by the model proposed in this Article. In a typical employment discrimination case, for example, the null hypothesis might be that the hiring decisions of the defendant reflected a random selection of men and women from the appropriate pool.

127 The 50% figure is based on the approximation of β suggested by Dawson, supra note 115. As noted, β cannot be determined precisely in situations in which there is an unknown distribution of possible true probabilities. See Kaye, supra note 112, at 14 n.7. Depending on the particular case, the value of β could be higher or lower. In most situations, however, the choice of $\alpha=0.05$ will result in β much greater than α.

the social disutility of wrongful acquittal. It is certainly not obvious that any values appropriate in civil litigation justify reaching this particular balance.

In order to determine a confidence level appropriate for use in constructing confidence intervals around probability estimates in civil litigation, we must examine a range of possible alternatives and their implications. Although such analyses have been undertaken in the narrow context of the hypotheses tested in discrimination litigation,[128] they have not been extended to civil litigation generally.

One possible solution is to use a confidence level that will minimize the likelihood that a defendant will lose if the true probability of the facts establishing his or her liability is less than 0.5. This choice has intuitive appeal because, if the factfinder had total information, the defendant would lose only if the probability with respect to the issue at hand exceeded 0.5. It would appear legitimate to assume that the legal system should seek to minimize the likelihood that a plaintiff for whom the relevant probability is actually less than 0.5 will be found to have met this burden. This alternative, however, does not withstand close scrutiny. The confidence level that most effectively minimizes the likelihood of a wrongful verdict in favor of a plaintiff is one hundred percent.[129] Such a high level of confidence would translate, in every case involving a factual dispute, into a very simple rule of decision: the plaintiff always loses. Thus, a system that achieves the goal of minimizing wrongful recoveries by plaintiffs by totally eliminating judgments for the plaintiff whenever the true (but unknown and unknowable) probability of the defendant's liability does not exceed 0.5 entails the high cost of producing the maximum number of incorrect judgments in favor of defendants.[130]

A second possible solution would be to choose a confidence level that most effectively minimizes the likelihood that a plaintiff will lose when the true probability in question exceeds 0.5. This alternative is the opposite of the first proposal, and it suffers from the same flaws. Using this solution, the rule of decision in all cases involving factual disputes would be equally simple: the plaintiff always wins.[131]

A third possibility would be to impose a preponderance test of sorts on the confidence interval by choosing a confidence level an infinitesimal amount in excess of fifty percent. At this confidence level, α, the risk of

---

128 See text accompanying notes 10-19, 122-24 supra.

129 See notes 106-07 and accompanying text supra.

130 This phenomenon follows from the inverse relationship between Type I error and Type II error. Here, where α, the risk of Type I error, is set at zero, β, the risk of Type II error, is at its maximum. See note 113 supra.

131 Although β, the risk of Type II error, would be zero, α, the risk of Type I error, would be at its maximum. See note 113 supra.

Type I error, would be a shade under 0.5, and β, the risk of Type II error, would probably be somewhat less.[132] This proposal is also inadequate, however. It is one thing to state that a plaintiff should win whenever the true (though unknown) probability of the facts in question exceeds 0.5. It is quite another to state that he or she should win whenever it is infinitesimally more likely than not that a defendant for whom the true probability is less than 0.5 would not reach the point estimate of probability. Even with total circumstantial information, a rule of decision that dictates a judgment for the plaintiff when the true probability is, say, 0.51 implies that we are willing to give the "wrong" verdict forty-nine percent of the time.[133]

Setting the confidence level at fifty percent means, of course, accepting a fifty percent chance of Type I error—that is, an even chance that a defendant for whom the true probability is 0.5 or less will nevertheless be found liable. Thus, a system that allows a plaintiff to prevail whenever the evidence supports the inference, at the fifty percent confidence level, that the probability exceeds 0.5 would tolerate not only the errors that would arise when the true probability was known, but also would include those arising from mistaken estimates of that probability. Such a system would give "incorrect" verdicts in favor of the plaintiff in a high percentage of cases in which the defendant should prevail. Furthermore, inasmuch as β, the risk of Type II error, would likely be less than α in these situations,[134] we would be making such mistakes proportionately more often against defendants than against plaintiffs without any policy justification for the imbalance.

Surprisingly, in the ideal world of normal distributions and random samples, this alternative would be the functional equivalent of the current probabilistic model, which incorrectly assumes that the factfinder determines the true probability of the defendant's liability. As Figure 3 illustrates, the lowest point in a one-tailed interval estimate at the fifty percent confidence level is the point estimate:[135]

---

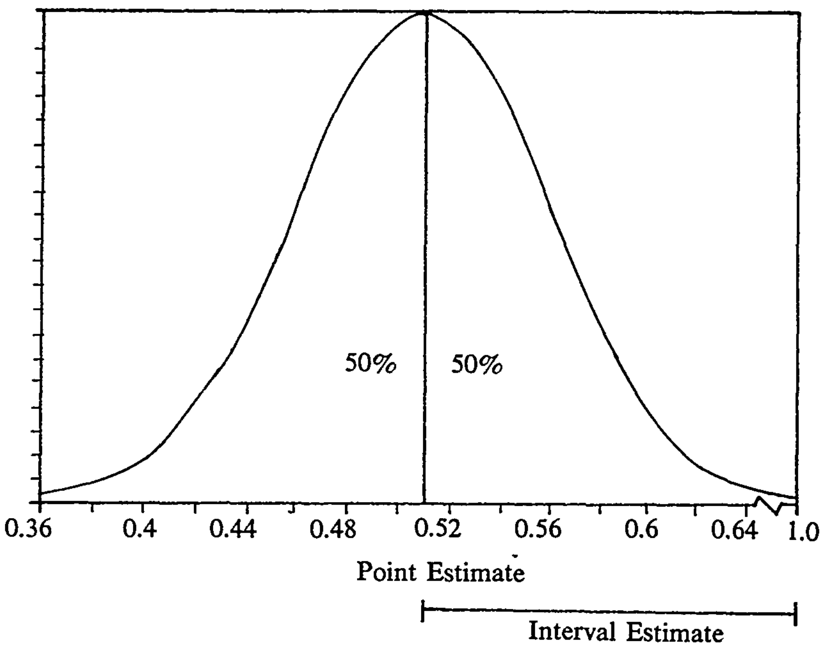
132 Although β is not directly determinable, see text accompanying notes 112-14 supra, because a β of roughly 0.5 has been shown in a prototypical case to be associated with an α of 0.05, and because there is an inverse relationship between α and β, the increase in α from 0.05 to 0.5 would cause a reduction in β from 0.5. See Dawson, supra note 115.

133 A 0.51 probability that the defendant actually committed the wrongful act is the same as a 0.49 probability that he did not. See Kaye, Naked Statistical Evidence, supra note 33, at 497.

134 See text accompanying notes 113-14 supra.

135 This is because, in such an ideal world, the probability curve straddles the point estimate evenly; therefore cutting off the lower 50% would leave only the area to the right of the point estimate. For a definition of one-tailed intervals, see text accompanying notes 88-89 supra.

FIGURE 3



Thus, a requirement that this entire interval exceed 0.5 is the same as a requirement that the point estimate exceed 0.5. The requirement that the factfinder's point estimate of the probability exceed 0.5 is, of course, the traditional probabilistic test.

Because the risk of false inculpation (Type I error) in this test is likely to be greater than the risk of false exculpation (Type II error), the test, in a sense, favors plaintiffs. Thus the traditional formulation of the preponderance of the evidence standard reflects an implicit and probably unintended bias in favor of one of the parties in litigation. This observation casts additional light on the inability of the traditional model to solve the problems posed by Cohen and Tribe. Recall that in each of these hypothetical situations the traditional model yielded a probability that was greater than 0.5 and therefore mandated a verdict for the plaintiff, despite arguments that the plaintiffs would not prevail in the legal system.[136] That problematic result can now be seen to flow from the assumption that a factfinder's point estimate of liability is the true probability (and the implicit adoption of the fifty percent confidence interval).

A fourth possibility would be to choose a confidence level at which β, rather than α, equals 0.5. Under this test, a defendant who should be

[136] See text accompanying notes 63-66 supra.

found liable would nevertheless have a fifty percent chance of prevailing. Such a model would have the same faults as the third alternative discussed except that it would disadvantage plaintiffs disproportionately, rather than defendants. At present, no policies have been articulated suggesting that we should prefer either of these alternatives.

A final possible solution would be to choose a confidence interval such that α equals β. Defendants would then lose when the true probability was under 0.5 as often as plaintiffs would lose when the probability exceeded 0.5. The costs incurred as a result of our inability to assess true probabilities because of our lack of complete information would fall evenly on plaintiffs and defendants. This alternative presents a significant practical difficulty because β, the risk of Type II error, cannot be determined directly.[137] To the extent that the new model is used heuristically rather than mechanically, however, this obstacle is not fatal.

Putting aside the practical difficulty, this fifth proposal is best suited to further the evenhanded policies underlying civil litigation.[138] Although in practice such a standard would not guarantee a "correct" result in any individual case, its overall effect would be to equalize the cost of "wrong" judgments so that the system as a whole would favor neither plaintiffs nor defendants.[139] Indeed, Professor Dawson has suggested this approach to assess allegations of discrimination.[140]

In sum, choosing the appropriate confidence level for the preponderance standard is not simple. Selecting the level that equalizes the risks of Type I and Type II errors has conceptual advantages, but this approach presents serious practical problems arising from our inability to determine β in the absence of information that is usually unavailable. Therefore, the best that we can do is to estimate the risk of Type II error and accept the inherent imprecision resulting from the assumptions used to make such estimates. Nevertheless, to the extent that the model serves a heuristic rather than a technical function, these problems do not detract from the model's usefulness.

137 See note 114 and accompanying text supra.

138 See Kaplan, supra note 19, at 1072.

139 Cf. Dawson, supra note 115, at 209 (making the same observation in the context of proving discrimination).

140 Id. However, neither Dawson nor the other legal statisticians who have done work in the discrimination area have noted the connection between the rules that they propose for determining the existence of discrimination and broader questions concerning probabilistic interpretations of the burden of persuasion.

III

Implications of the New Model

A. *Allocation of the Burden of Persuasion*

The approach that this Article proposes for understanding the preponderance burden of persuasion in probabilistic terms is essentially a redefinition of the quantum of evidence that the party bearing the burden must present in order to prevail. Viewed in this way, the model has important implications for the allocation of burdens of persuasion among the parties to litigation. Under the traditional formulation of the preponderance burden, which is described as the burden of demonstrating that the true probability of the event in question exceeds 0.5,[141] the allocation of the burden would not have very much significance. If that formulation were accurate, the only circumstance in which the allocation would determine the result of a case would be a situation in which neither party could meet the burden—that is, when the probability in question equals exactly 0.5. In such a case, the party bearing the burden of persuasion would lose. In all other situations, the result would not be affected by the placement of the burden of persuasion. Because in any case in which any evidence has been introduced the chance that the probability equals exactly 0.5, and not 0.50001 or 0.49999, is minuscule, the current probabilistic formulation of the burden of persuasion implies that allocating that burden should not be an important task. Yet any student of evidence or civil procedure knows that courts and litigants rarely consider the burden of persuasion to be a minor issue. Indeed, virtually every area of the law is replete with cases debating the allocation of the burden of persuasion.[142]

Conceiving of the burden of persuasion as presented in this Article, on the other hand, sheds light on why burden allocation is taken so seriously. This new model clarifies the practical content within the concept of equipoise. Situations in equipoise—in which neither party can meet the burden of persuasion—would include not only situations where the true probability is exactly 0.5 but also all situations in which the interval estimate of the probability of the facts supporting liability straddles 0.5. In any such case, the evidence provided by the parties would be insuffi-

---

141 See text accompanying note 61 supra.

142 See, e.g., Russell v. Barnes Found., 52 F. Supp. 827 (E.D. Pa. 1943), aff'd, 143 F.2d 871 (3d Cir.), cert. denied, 323 U.S. 771 (1944) (anticipatory breach of employment contract); Crillo v. Curtola, 91 Cal. App. 2d 263, 204 P.2d 941 (1949) (breach of employment contract); Quillen v. Betts, 48 Del. 93, 98 A.2d 770 (1953) (in motion to recall bail after indictment, burden placed on defendant to rebut presumption of guilt); Nelson v. Oversmith, 69 Idaho 1, 201 P.2d 747 (1949) (to recover damages, burden on plaintiff to show negligence); Milhollin v. Adams, 66 Ind. App. 376, 115 N.E. 803 (1917) (wrongful discharge). See also 9 J. Wigmore, supra note 57, §§ 2485-2489.

cient to allow the factfinder to state with sufficient confidence that the probability that the facts support either party's position exceeds 0.5. Accordingly, in these cases it is important to determine which party will suffer for that mutual inability. Under the reformulated definition of the quantum of the burden of persuasion, allocation of the burden determines who will lose when the factfinder cannot determine with the requisite amount of confidence on which side of 0.5 the true probability lies.[143]

B. *Harmonizing Opinions that Reject Probabilistic Analysis*

This new formulation of the burden of persuasion also helps to harmonize probabilistic legal analysis with judicial opinions previously considered hostile to the use of such techniques. The case most often cited for the proposition that a probabilistic interpretation of the preponderance standard is inappropriate is *Sargent v. Massachusetts Accident Co.*[144] In this opinion, the Supreme Judicial Court of Massachusetts stated:

> It has been held not enough that mathematically the chances somewhat favor a proposition to be proved; for example, the fact that colored automobiles made in the current year outnumber black ones would not warrant a finding that an undescribed automobile of the current year is colored and not black, nor would the fact that only a minority of men die of cancer warrant a finding that a particular man did not die of cancer. The weight or ponderance of evidence is its power to convince the tribunal which has the determination of the fact, of the actual truth of the proposition to be proved. After the evidence has been weighed, that proposition is proved by a preponderance of the evidence if it is made to appear more likely or probable in the sense that actual belief in its truth, derived from the evidence, exists in the mind or minds of the tribunal notwithstanding any doubts that may still linger there.[145]

Five years later the same court relied upon this dictum as the basis for rejecting statistical evidence in *Smith v. Rapid Transit, Inc.*,[146] the case that inspired Professor Tribe's bus hypothetical.

At first glance, these cases seem squarely to reject explicit probabilistic interpretations of the preponderance standard. Indeed, *Sargent* has frequently been cited in support of that proposition.[147] Yet a close read-

---

143 My colleague D. Michael Risinger has long held a similar view. He contends that factfinders who consider a case to be in equipoise do not in fact believe that they have measured precisely the relevant probability to be exactly 0.5; rather, their view of the probability tends to oscillate between values on either side of 0.5 and they cannot confidently fix it on one side or the other.

144 307 Mass. 246, 29 N.E.2d 825 (1940).

145 Id. at 250, 29 N.E.2d at 827 (citations omitted).

146 317 Mass. 469, 58 N.E.2d 754 (1945).

147 See, e.g., Callen, supra note 33, at 37; Kaye, The Paradox of the Gatecrashers, supra note 33, at 103-11; Winter, supra note 61, at 338. See also Stepakoff v. Kantar, 393 Mass. 836,

ing of the *Sargent* court's statement can support the related, but fundamentally different, conclusion that the court actually was rejecting a particular probabilistic interpretation of the preponderance of the evidence standard that it viewed as mistaken—the one that treats point estimates as true values.

Indeed, much of the court's language is not necessarily inconsistent with the confidence level interpretation of the preponderance of the evidence standard. In particular, the two examples the *Sargent* court provides—the probability that a randomly selected automobile will be colored or that a man who dies had cancer—resemble the cases of "naked statistical evidence" posited by Cohen and Tribe.[148] Once the sample statistic nature of probability estimates is recognized, the court's examples easily can be harmonized with probabilistic formulations of the preponderance standard.

Although the *Sargent* court's references to the need to persuade the tribunal of the "actual truth of the proposition to be proved"[149] and the need for the tribunal to have an "actual belief in its truth"[150] are much harder to reconcile with the probabilistic view, these references can be understood as unfocused attempts to describe the concept of confidence. Thus, these cases may display judicial hostility more toward a particular conception of probabilistic proof than to the general appropriateness of probability in the legal context. The Massachusetts court may have been expressing, at an intuitive level, a view that is consistent with the ideas presented in this Article.

C. *Implications for Other Burdens of Persuasion*

This Article has focused on the burden of persuasion most commonly applied in civil lawsuits—the preponderance of the evidence standard. Nevertheless, the analysis set forth is relevant to other burdens of persuasion as well. In many civil lawsuits involving fraud or other quasi-criminal conduct, the burden of persuasion requires the plaintiff to demonstrate the relevant fact or issue by "clear and convincing evidence."[151] In criminal cases, courts apply the even stricter standard of "proof beyond a reasonable doubt."[152]

—, 473 N.E.2d 1131, 1136 (1985) (judge's charge quoting *Sargent* language on appeal); King's Case, 352 Mass. 488, 491-92, 225 N.E.2d 900, 902 (1967) (favorably referring to *Sargent*).

[148] See Kaye, Naked Statistical Evidence, supra note 33, at 487-90. "Naked statistical evidence" refers to cases in which there is *only* quantitative evidence.

[149] 307 Mass. at 250, 29 N.E.2d at 827.

[150] Id.

[151] C. McCormick, supra note 57, § 340, at 959; 9 J. Wigmore, supra note 57, § 2498, at 424-27.

[152] C. McCormick, supra note 57, § 341, at 962; 9 J. Wigmore, supra note 57, § 2497, at 404-15.

Prior theorists have assumed that the probabilistic formulations of these more stringent burdens differ from the preponderance standard in requiring that the probability of the facts supporting liability or guilt exceed a threshold higher than 0.5.[153] In the criminal area, much effort has been expended in debating the precise level of that threshold probability.[154] This Article has demonstrated, however, that there are two variables in the probabilistic formulation of burdens of persuasion—the threshold probability and the level of confidence that is used in constructing the interval estimate.

Accordingly, the more stringent burdens of persuasion can differ from the preponderance burden in one or both of two ways. Either the threshold may be increased (which has been the presumption of current theories),[155] or the confidence level may be increased, or both. For example, the difference between the clear and convincing standard and the preponderance of the evidence standard could be that the former has a higher threshold probability (which is the only possibility in the current model). The difference, however, also could be that the clear and convincing standard requires the factfinder to use a higher level of confidence in constructing the interval estimate. The burden then would be described as demonstrating at the new, higher confidence level that the estimated probability of the facts supporting liability exceeds 0.5. If the confidence required is increased substantially, even though the 0.5 threshold value remains unchanged, the plaintiff's burden would be significantly more difficult to meet. By allowing for the interaction of two variables, the probabilistic model identified in this Article will enable more sophisticated analysis of burdens of proof generally.[156]

CONCLUSION

Current probabilistic models of burdens of proof have encountered significant difficulty in attempting to account for discrepancies between the results that they suggest and those that the legal system regularly reaches. These models suggest that the burden has been met in many cases in which courts and juries would not so hold—particularly in cases in which the factfinder has relatively little information. Proponents of the current probabilistic models have attempted to explain these discrepancies by arguing that they result from the presence of other values

153 See, e.g., Kaplan, supra note 19, at 1072.

154 See, e.g., id. at 1073-77.

155 See, e.g., United States v. Fatico, 458 F. Supp. 388 (E.D.N.Y. 1978), aff'd, 603 F.2d 1053 (2d Cir. 1979), cert. denied, 444 U.S. 1073 (1980).

156 Furthermore, this author will demonstrate in a forthcoming work how the new model can illuminate the process by which courts review findings of fact and the standards applied in such reviews.

within the legal system that take precedence over probabilistic factfinding.

This Article contends that these discrepancies are not the result of external values but rather follow from a fundamental error in the models. The current theories assume that the legal system can and does determine the true probabilities of disputed legal facts. They then define the burden of persuasion as the burden of demonstrating that the probability of the facts supporting liability exceeds a particular threshold value.

This Article explains that the legal system can only estimate probabilities. Once it is recognized that forensically determined probabilities are only estimates, it becomes necessary to develop a new probabilistic model of the burden of persuasion. The model developed in the Article requires that, in order to satisfy the preponderance of the evidence standard, not only must the factfinder's best guess, or point estimate, of the probability of the facts supporting liability exceed 0.5, but the factfinder also must be confident that the true probability of liability in fact exceeds 0.5. In statistical terms, the entire interval estimate, or confidence interval, of the probability must exceed 0.5 at a given confidence level.

This model offers a more accurate, comprehensive concept of forensically determined probabilities than do current models of probability. By taking into account the role of confidence as well as the factfinder's "best guess," this new model avoids many of the difficulties encountered by the previous models and clarifies our understanding of equipoise and of more stringent burdens of persuasion.