1979

Book Review

Elaine W. Shoben
University of Nevada, Las Vegas -- William S. Boyd School of Law

BOOK REVIEW

QUANTITATIVE METHODS IN LAW: STUDIES IN THE APPLICATION OF MATHEMATICAL PROBABILITY AND STATISTICS TO LEGAL PROBLEMS. *By Michael O. Finkelstein.* The Free Press. New York. 1978. Pp. xi, 318. $17.95

*Reviewed by Elaine W. Shoben**

I. INTRODUCTION

Many cases require legal factfinders to confront questions that are probabilistic in nature. What is the likelihood that blacks would be underrepresented on jury panels to a certain degree if there was no intentional exclusion? How great is the chance that a primary election result would be different if the invalid votes had not been cast? What is the probability that a driver neglected to set the handbrake carefully if a parked car rolls down a hill?

Most individuals are remarkably poor estimators of frequencies. People consistently overestimate the number of airplane crashes and underestimate the number of pathways between two points.[1] Consequently, when they are asked to make an intuitive estimate of the probability of a particular event given its relative frequency, the answers necessarily reflect the underlying misestimates. Similarly, when subjective judgments about numerical processes are made, patterns of objective error are made. Individuals tend to judge the product of $5 \times 4 \times 3 \times 2$ to be larger than the product of $2 \times 3 \times 4 \times 5$, although the actual calculations would prove them to be objectively identical.[2]

These fallacies of intuitive estimates may be merely inconvenient in a card game, but in important legal decisions these subjective answers provide very poor guides. When the fact-finding process in legal adjudications involves assessments of probabilities or judgments on numerical evidence, reliance on rough subjective estimates can produce unfair results. Judges, administrative bodies, civil and criminal juries are often asked to make such judgments with little or no guidance on how to proceed.

Quantitative Methods in Law[3] represents the efforts of one legal scholar to apply mathematical probability and statistics to the solution of a wide range of legal problems. Michael O. Finkelstein has republished in book form a collection of his articles, beginning with his most famous and most widely cited: the application of mathematical probability to jury discrimi-

---

* Associate Professor of Law, University of Illinois, College of Law. A.B. 1970, Barnard College; J.D. 1974, University of California, Hastings College of the Law.

[1] *See* Tversky and Kahneman, Availability: A Heuristic for Judging Frequency and Probability, 5 Cognitive Psych. 207, 213, 230 (1973) [hereinafter cited as Tversky].

[2] *See* discussion of these examples and other related ones at text accompanying notes 55-60 *infra.*

[3] M. Finkelstein, Quantitative Methods in Law: Studies in the Application of Mathematical Probability and Statistics to Legal Problems (1978).

nation cases.[4] After leading the reader through a series of fascinating applications of statistical problem solving to an impressively wide range of legal situations, the book concludes with the final words of one of the most engaging battles among legal scholars in recent years: the exchange between Michael Finkelstein and Laurence Tribe on the use of Bayes' theorem in a criminal trial to assist the jury in integrating probabilistic evidence with nonnumerical testimony.[5]

The ideas in Finkelstein's book can be divided into basically two categories which can be denominated "responsive" and "innovative" uses of quantitative methods in law. The first, the responsive uses, consists of those chapters that propose the use of more precise calculation to answer legal questions already well defined in probabilistic terms. Most prominent in this category is the use of probability theory in jury discrimination cases. The second category, innovative uses, includes those chapters which suggest reformulation of the relevant factual questions in a manner compatible with Finkelstein's more scientific approach to inquiry. This category includes the author's excursions into the areas of weighted voting,[6] economic concentration[7] and solvency controls.[8] Two miscellaneous chapters, those on wrongful death damages[9] and guilty pleas,[10] subject some common assumptions on those subjects to empirical scrutiny. All of these uses can be called "quantitative methods," as the title of the book broadly groups them. The persuasiveness of each of Finkelstein's models, however, depends on its function. Responsive uses—models suggesting quantitative methods to replace intuitive responses to probabilistic legal

[4] *Id.*, Chapter 2 ("Jury Discrimination"), *originally appearing as* Finkelstein, The Application of Statistical Decision Theory to the Jury Discrimination Cases, 80 Harv. L. Rev. 338 (1966).

[5] The original article, now Chapter 3 ("Two Cases in Evidence"), of M. Finkelstein, *supra* note 3, was: Finkelstein & Fairley, A Bayesian Approach to Identification Evidence, 83 Harv. L. Rev. 489 (1970). The critique of his approach by Professor Tribe then appeared a year later. Tribe, Trial by Mathematics: Precision and Ritual in the Legal Process, 84 Harv. L. Rev. 1329 (1971). Finkelstein's reply appeared then as: Finkelstein & Fairley, A Comment on "Trial by Mathematics," 84 Harv. L. Rev. 1801 (1971). Finally, Tribe's response to the reply appeared as: Tribe, A Further Critique of Mathematical Proof, 84 Harv. L. Rev. 1810 (1971). The reply and response both are republished as the Appendix in M. Finkelstein, *supra* note 3, at 288. For a discussion of the use of Bayes' theorem in integrating probabilistic evidence with nonnumerical testimony, see text accompanying notes 63-64 *infra*.

[6] M. Finkelstein, *supra* note 3, Chapter 4 ("Voting"), *originally appearing as:* Finkelstein & Robbins, Mathematical Probability in Election Challenges, 73 Colum. L. Rev. 241 (1973).

[7] *Id.*, Chapter 5 ("Economic Concentration"), *originally appearing as:* Finkelstein & Friedberg, The Application of an Entropy Theory of Concentration to the Clayton Act, 76 Yale L.J. 677 (1967). A response was written to follow this article as: Stigler, Comment, 76 Yale L.J. 718 (1967). Finkelstein's brief reply appears following the comment: Finkelstein & Friedberg, Reply to Professor Stigler, 76 Yale L.J. 721 (1967).

[8] *Id.*, Chapter 6 ("Solvency Controls"), *originally appearing as:* Finkelstein, The Use of Risk Theory in Framing Solvency Controls for Nonlife Insurance Companies, 119 U. Pa. L. Rev. 730 (1971).

[9] *Id.*, Chapter 8 ("Compensation for Wrongful Death"), *originally appearing as:* Finkelstein, Pickrel & Glasser, The Death of Children: A Nonparametric Statistical Analysis of Compensation for Anguish, 74 Colum. L. Rev. 884 (1974).

[10] *Id.*, Chapter 9 ("Guilty Pleas"), *originally appearing as:* Finkelstein, A Statistical Analysis of Guilty Plea Practices in the Federal Courts, 89 Harv. L. Rev. 293 (1975).

questions—should be more readily acceptable than innovative uses that suggest redefining the legal approaches to some particular problems.

These distinctions in uses of quantitative methods regrettably are not salient throughout Finkelstein's book. It is clearly relevant to the reader's assessment of the merits of each technique to be certain whether the author is offering new ways of answering old questions or whether he is asking new questions. Those chapters concerned with innovative uses can be evaluated best only in the context of the fields they address. Regardless of whether they offer definitive new solutions to the problems they consider, they provide a fresh perspective to old issues and help to shake one out of a "fixed mental set" on how to solve particular legal problems. As such, those discussions are provocative ones for all readers.[11]

This review will focus initially on Finkelstein's responsive use of quantitative methods of analysis: quantitative methods to give greater precision in situations where courts presently rely on intuitive assessments to answer probabilistic questions. The reason for focusing on the responsive uses is not only that the diversity of the innovative models makes them unsuitable for single review, but more importantly, that Finkelstein's crusade against the use of simple intuition to answer factual questions of probability is an important contribution to the law. These uses should not be confused with his other creative, but controversial, proposals for quantitative methods in the law. This review next will address one of Finkelstein's innovative models: his proposed use of Bayes' theorem in criminal trials.[12] Some of the problems with subjective misestimation that inspired his model may cause his solution to compound the factfinders' misestimation of probabilities. Throughout both discussions this review refers to the work of two prominent psychologists who have explored the nature of subjective misestimates of probabilities when individuals answer such questions intuitively. Their work underscores the challenge to the law to avoid fact-finding rules which comport with intuitive estimates but not with more objective calculations.

II. Erroneous Probability Assessments—The Need for Precision

Some legal questions require as part of the fact-finding process an assessment of a particular probability. Such analysis has figured prominently in employment discrimination cases under title VII of the Civil Rights Act of 1964,[13] as well as in the areas of jury discrimination and irregular elections. A typical kind of legal fact-finding probability question is: what is the probability of obtaining a sample group, such as a jury, of a

[11] For a review of some of Finkelstein's innovative models, see Brilmayer & Kornhauser, Review: Quantitative Methods and Legal Decisions, 46 U. Chi. L. Rev. 116 (1978).

[12] *See* note 5 *supra*.

[13] *See* Hallock, Numbers Game—The Use and Misuse of Statistics in Civil Rights Litigation, 23 Vill. L. Rev. 5 (1977); Shoben, Probing the Discriminatory Effects of Employee Selection Procedures with Disparate Impact Analysis Under Title VII, 56 Tex. L. Rev. 1 (1977); Note, Beyond the Prima Facie Case in Employment Discrimination Law: Statistical Proof and Rebuttal, 89 Harv. L. Rev. 387 (1975).

particular composition from an unbiased selection from a given population? If the court relies on intuition to answer such a question, serious errors in the assessment are likely because there are systematic fallacies in intuition not encountered in appropriate mathematical calculation. The likelihood and gravity of these errors are explored in the following sections, followed by a defense of Finkelstein's methods of calculation.

A. *Jury Composition: Probability of Particular Rácial Underrepresentation*

One of the most prominent of the probabilistic legal problems is whether a particular underrepresentation of a racial minority on juries reflects an intentional exclusion of that group or reflects a likely outcome of a random selection process. The relevant legal question, as gradually formulated by the Supreme Court,[14] is whether this deviation from the "expected" percentage representation is sufficiently unlikely to happen by chance alone that the inference of intentional discrimination is supported.[15] The Supreme Court has even suggested in dicta that if the result is extreme, the numerical discrepancy alone may be sufficient to infer intentional racial discrimination in jury cases.[16]

The fact-finding problem then becomes one of assessing the rareness of a particular result. In the absence of guidance provided by counsel,[17] courts have had to rely on intuitive estimates of likelihood. The difficulty with reliance on intuition is that individuals tend to make systematic errors in estimating probabilities. Psychologists Daniel Kahneman and Amos Tversky have studied extensively the nature of subjective probability judgments and conclude that "people do not follow the principles of probability theory in judging the likelihood of uncertain events"[18] because the laws of chance are not intuitively apparent. They note further that "the deviations of subjective from objective probability seem reliable, systematic and difficult to eliminate. Apparently, people replace the laws

---

[14] *See* Castaneda v. Partida, 430 U.S. 482 (1977); Alexander v. Louisiana, 405 U.S. 625 (1972); Turner v. Fouche, 396 U.S. 346 (1970); Whitus v. Georgia, 385 U.S. 545 (1967); Cassell v. Texas 339 U.S. 282 (1950).

[15] Castaneda v. Partida, 430 U.S. 482 (1977), clarified that in order to establish a prima facie case of discrimination in the selection of jurors, a defendant must show (1) that the group discriminated against is a distinct class, (2) that the group is underrepresented over a substantial period of time, and (3) that the selection process is not racially neutral or is susceptible to be used as a tool of discrimination. *Id.* at 494-95. This discussion goes to the proof of part (2) of these elements. *See also* Rose v. Mitchell, 99 S. Ct. 2993, 3005 (1979).

[16] Castaneda v. Partida, 430 U.S. 482, 493-94 (1977) (citing Washington v. Davis, 426 U.S. 229, 241 (1976), and Arlington Heights v. Metropolitan Hous. Dev. Corp., 429 U.S. 252, 264-65 (1977)). The distinction made by the court in *Castaneda* is that although an official act is not unconstitutional *solely* because it has a disproportionate impact, nonetheless a clear pattern of intentional discrimination can appear from the statistics themselves. 430 U.S. at 493-94.

[17] For a trial judge's lament on the failure of counsel to do an adequate job of interpreting statistics for the court in any manner "other than the crudest fashion," see Garrett v. R. J. Reynolds Indus., Inc., 81 F.R.D. 25, 32-33 & n.9 (M.D.N.C. 1978).

[18] Kahneman & Tversky, Subjective Probability: A Judgment of Representativeness, 3 Cognitive Psych. 430, 431 (1972) [hereinafter cited as Kahneman].

of chance by heuristics, which sometimes yield reasonable estimates and quite often do not."[19]

What, for example, is the likelihood of obtaining only four percent blacks on juries when blacks constitute twenty percent of the registered voters eligible for jury duty? Is this discrepancy sufficiently large to reflect a systematic exclusion of blacks, or is it a difference that could easily happen by chance alone? The most common approach taken by judges in such cases has been to subtract the two percentages and intuit whether this difference is large. The Supreme Court used this approach in *Swain v. Alabama*[20] in 1965, and concluded that a difference of only ten percentage points was not sufficient to amount to a constitutional violation. Although the Supreme Court has used other assessments since *Swain*,[21] that method continues to be used in some circuits.[22]

The difficulty with the *Swain* method of assessing the discrepancy between the expected and the observed number of black jurors is that it obtains inconsistent results among cases. This "percentage point discrepancy" approach fails to take into account either the size of the sample of jurors in question or the difference between a large or small representation of eligible blacks. A difference of ten percentage points is much rarer in a large sample than in a small one.[23] Similarly, a difference of ten percentage points is much rarer when the minority group is a small percentage of the eligible voters than it is when the minority representation approaches one half.[24]

The studies of Kahneman and Tversky[25] explain why factfinders naturally would tend to adopt a strategy such as the misleading ten percentage point discrepancy approach. Their research indicates that individuals tend to ignore crucial differences in sample size when assessing intuitively the likelihood of particular results. Objectively, however, sample size is an indispensable piece of information; eighty percent heads when flipping a coin is not an unusual result if there are only five flips, but it is an extremely rare result if there are a million flips. That

---

[19] *Id.*

[20] Swain v. Alabama, 380 U.S. 202, 208-09 (1965).

[21] *See* note 15 *supra*.

[22] *See, e.g.*, Ross v. Wyrick, 581 F.2d 172 (8th Cir. 1978); United States v. Brady, 579 F.2d 1121 (9th Cir. 1978), *cert. denied*, 99 S. Ct. 849 (1979); Berry v. Cooper, 577 F.2d 322 (5th Cir. 1978); United States v. Newman, 549 F.2d 240 (2d Cir. 1977); Blackwell v. Thomas, 476 F.2d 443 (4th Cir. 1973).

[23] *See* text accompanying notes 25-27 *infra*.

[24] Assume, for example, minority representation is only 10% in the population and out of a sample of 100 people from this population there are no minorities. The difference is 10 percentage points. The probability of such a result this extreme or more so is less than 1 in 500.

On the other hand, assume that minority representation in the population is 40% and out of a sample of 100 from this population, 30% or 30 of the minority are found. The difference in percentages is again 10 but the probability of a result this extreme or more so by chance alone is approximately 1 in 17.

[25] Kahneman, *supra* note 18, at 437-45.

property of sampling is easily agreed upon in principle, but apparently individuals consistently ignore it when they are asked to make estimates.

In one study by Kahneman and Tversky, for example, people were asked to consider the numbers of baby boys and girls born daily in small, medium and large hospitals. In the small hospital 10 babies are born a day, so one would expect an average of 5 boys a day. In the medium sized hospital 100 babies are born a day, so one would expect an average of 50 boys. In the large hospital, 1000 babies are born a day, making an expected average of 500. The researchers told individuals about one of these hypothetical hospitals and then asked them to estimate how many days a year one would expect to find various deviations from the expected average number of boys. In the medium sized hospital situation, for example, the questions would be: on what percentage of days in a year will the number of boys among 100 babies be up to 5? 5 to 15? 15 to 25? and so forth. The individuals were thus asked to generate distributions of what they estimated would be the relative frequency of particular deviations from the expected average number of boys overall.

The rather startling result of this study was that people tend to estimate the same degree of deviation regardless of the size of the hospital. For example, 5-15 percent boys was deemed as likely to happen on the same number of days in a small hospital with 10 babies a day—1 boy out of 10 babies—as in a large hospital with 1000 babies a day—50-150 boys out of 1000 babies. Objectively this difference in sample size—10 versus 1000—is crucial because in the large sample there is a greater opportunity for daily averaging; in a large hospital there may be more of one sex born in any given hour, but by the end of the day the 1000 babies will probably be very close to 500 boys and 500 girls. When only 10 babies are born a day, however, there is less chance to average out by the end of the day, so on any given day there may be greater deviation from the expected average of 5.

The researchers attributed this failure of individuals to account subjectively for sample size to an overreliance on the "representativeness" of the sample result. There is a universal mental curve of an imagined typical degree of variance from the average, and that representation assumes as much variance in large samples as in small ones.[26] Thus, individuals erroneously estimated that 200 boys and 800 girls would happen as often in the large hospital as 2 boys and 8 girls would happen in the small hospital.[27]

In the jury discrimination cases the same fallacy would lead most people to conclude that if a minority group were represented on juries to a degree within 10 percentage points of their availability—the expected

---

[26] *Id.* at 444.

[27] The probability of the 200 and 800 split, or one even more extreme, at the large hospital is far less than one in a billion. At the small hospital, however, the probability of a split as extreme as 2 and 8 or more so is about 1 in 9.

average—the result would not be very rare. In objective terms the difference in percentages is insufficient information on which to make a judgment: sample size makes the key difference. When the sample size is large, as it often is in jury discrimination cases, a difference of a few percentage points may be very unlikely to happen by chance alone.

Perceiving the need for more objective measurement of the degree to which a minority is underrepresented on juries, Finkelstein suggests the use of a straightforward calculation of the binomial probability involved in such cases. His explanation of this calculation originally appeared in a 1966 article,[28] and it is now reprinted prominently in *Quantitative Methods in Law.* The article has been widely cited by federal courts at all levels.[29] It remains an outstanding contribution to the evidentiary law of jury discrimination cases.

The analysis Finkelstein employs is based on hypothesis testing: what is the probability of obtaining an underrepresentation of a minority group on juries as extreme or more so than the one in question if it is true that persons are chosen from the available pool without bias?[30] If the probability of the result happening by chance alone is very low—below a chosen critical level, often one in twenty or 0.05—then the result is deemed to be statistically significant. The value of this approach is that it provides a direct answer to the question of how rare is the underrepresentation if the selection process is truly random. A court may then use this information as it sees fit in determining whether there has been unlawful discrimination against the minority group.

Finkelstein's use of the statistical significance approach has drawn criticism in one scholarly article[31] on basically two grounds: that it is difficult to visualize the test, and that jury discrimination cases are improperly treated as sampling situations.[32] The first objection, the difficulty in apply-

---

[28] Finkelstein, *supra* note 4.

[29] *See, e.g.,* Castaneda v. Partida, 430 U.S. 482, 496 n.17 (1977); Whitus v. Georgia, 385 U.S. 545, 552 n.2 (1967); Smith v. Yeager, 465 F.2d 272, 278 n.16 (3d Cir.), *cert. denied,* 409 U.S. 1076 (1972); Gibson v. Blair, 467 F.2d 842, 844 n.1 (5th Cir. 1972); United States v. Butera, 420 F.2d 564, 568-69 n.11 (1st Cir. 1970); Witcher v. Peyton, 405 F.2d 725, 726 n.3 (4th Cir. 1969); Goins v. Allgood, 391 F.2d 692, 699 n.6 (5th Cir. 1968).

[30] For additional information on the nature of hypothesis testing in applied statistics, see H. Blalock, Social Statistics 151-66 (2d ed. 1972); L. Horowitz, Elements of Statistics for Psychology and Education 157-71 (1974).

[31] Kairys, Kadane & Lehoczky, Jury Representativeness: A Mandate for Multiple Source Lists, 65 Cal. L. Rev. 776, 794 (1977) [hereinafter cited as Kairys].

[32] *Id.* The authors' view that the jury cases are not properly characterized as sampling situations causes them to object that sample size differences play too large a role in Finkelstein's method. They state that the use of significance tests are inappropriate because:

> The source is not the result of drawing from a population but a list or combination of lists compiled for other purposes and by other means, and the pool is the result of a drawing from the source . . . . [T]he relevant question is whether the source or pool provides an adequate cross section of the community, not whether it is a random sample of the community.

Id. at n.101. As such, they advocate the use of the comparative disparity test to provide a standard for comparing the similarity of two separate distributions rather than the use of ing the concept, overstates the complexity of the calculation[33] and understates the abilities of the judiciary. Whatever method of analysis is used, accuracy and reliability must be considered more compelling than simplicity.

More importantly, the critics assert that the statistical significance test is inappropriate in jury cases because the appropriate legal question to ask is whether the source pool or juries are *representative* of the population, not whether there has been any purposeful *discrimination* in the selection process.[34] To date, the Supreme Court has addressed the constitutional question as one of discrimination, not representation, as the critics fully acknowledge.[35] Their alternative analysis must therefore be viewed as an innovative suggestion on how to approach the formulation of the legal issue and not as a responsive approach to the legal question as it is presently conceived. Discrimination questions involve a sampling process—drawing smaller groups from a larger one supposedly without bias. As such, Finkelstein's method is superior to other available alternatives.[36]

Finkelstein's model. Finkelstein tests whether the jury source or pool was randomly—without bias—drawn as a sample from the larger relevant population. Thus the authors' objection to the effect of sample size is inextricably tied to their objection to a sampling model. For further discussion of which model most directly addresses the legal issue see, notes 34-36 *infra*.

[33] *See* M. Finkelstein, *supra* note 3, at 34-35.

[34] Kairys, *supra* note 31, at 779-83.

[35] *Id.* at 785-86.

[36] Kairys lists four approaches to measuring representativeness, not discrimination. The first is the absolute disparity rule used in *Swain,* discussed at notes 20-24 and accompanying text *supra.* Kairys agrees that the absolute disparity rule is inadequate both legally and mathematically. Kairys, *supra* note 31, at 789-90, 793-94. They also reject the statistical significance test advocated by Finkelstein for the reasons discussed in notes 32-35 *supra;* basically, they find it an inadequate measure of *representativeness.* The other two approaches they list are the "comparative disparity" standard and the "proportion of eligibles" standard. They note that these two approaches are directly related mathematically and logically. Of these two, the authors prefer the comparative disparity measure, which shows the reduced probability of a group member's serving on a jury. Kairys, *supra* note 31, at 790-92.

Although it is beyond the scope of this review to evaluate in detail the comparative disparity standard as a measure of *discrimination,* not representation, its shortcomings lie in basically two areas: its failure to account for sample size differences which are crucial differences in sampling models and its failure to handle differences in the magnitude of the underrepresentation. As an illustration of the magnitude problem, consider two equally sized towns with different racial compositions: Town A is 10% black but Town B is 80% black. Assume that 1,000 people serve as jurors in both towns of equal sample size. In Town A, 8.5%, or 85 of the jurors are black instead of the expected 10%, or 100. In Town B, 69%, or 690 of the jurors are black instead of the expected 80%, or 800. The difference in the magnitude of underrepresentation is much greater in Town B—a difference of one or two blacks fewer than expected on the average on every jury panel. Viewed as a discrimination problem using Finkelstein's approach, Town A's underrepresentation would be found likely to have happened by chance alone, with about one chance in eight of this result happening without bias in the selection; Town B's underrepresentation of blacks, however, is highly unlikely to happen by chance alone, with a probability far less than one in a million. Using the comparative disparity approach with a 15% allowable underrepresentation, as advocated by Kairys, results in opposite conclusions: Town A would be found to be unconstitutionally underrepresenting blacks and Town B would not. Whether this result is defensible under a representation model remains to be settled by the courts, but it should not be considered a

B. *Elections: Irregular Votes*

Another legal question that has a probability component concerns whether an election should be overturned when the presence of irregular votes is established. Finkelstein addressed this problem in part of his 1973 article on voting,[37] now appearing as chapter four in his book. His discussion centers on a series of New York cases where the number of irregular votes exceeded the margin of victory of the winning candidate in a primary election.[38] The legal problem was whether the presence of the irregular votes was likely to have affected the outcome of the primary, thus necessitating a new election.

The standard articulated by the New York Court of Appeals for deciding whether to invalidate the election was if the "irregularities are sufficiently large in number to establish the probability that the result would be changed . . . ."[39] Finkelstein notes that the standard was then implemented by an intuitive assessment of the probability of a different result, and he suggests instead that the probability should be calculated with more precision.[40]

The difficulty in estimating the likelihood of an election reversal if the irregular votes are removed is similar to the problems in assessing the cases involving the underrepresentation of a minority group on juries. If one starts with the assumption that the votes are expected to be split equally between the two neck-and-neck primary candidates—an assumption which clearly is not warranted if there is any evidence of fraud in the irregular votes—then the question becomes: how likely is it that a large enough deviation from the expected split of the votes would overtake the winner's margin of victory?

Finkelstein characterizes the problem as similar to drawing colored marbles out of an urn. The different colors reflect votes for the different candidates. Some of the marbles have been put in the urn erroneously—the irregular votes—and must be withdrawn. There is no way of knowing the exact numbers of each color—or irregular votes—that must be withdrawn, so the process of withdrawing them is governed by chance, unless there is reason to suspect a pattern in the votes cast irregularly. If the

good approach to discrimination questions under the present law. These objections to the comparative disparity model for discrimination analysis—its failure to account for sample size and its problem handling different magnitudes of difference—are similar to the problems encountered with another arbitrary rule of thumb for assessing discrimination, the "four-fifths rule" used by the Uniform Guidelines on Employee Selection Procedures. *See, e.g.*, Shoben, Differential Pass-Fail Rates in Employment Testing: Statistical Proof Under Title VII, 91 Harv. L. Rev. 793 (1978). Unfortunately the comparative disparity approach has appeared occasionally in the discrimination context. *See* Foster v. Sparks, 506 F.2d 805 (5th Cir. 1975) (Appendix at 811-37).

[37] *See* note 6 *supra*.

[38] DeMartini v. Power, 27 N.Y.2d 149, 262 N.E.2d 857, 314 N.Y.S.2d 609 (1970); Santucci v. Power, 25 N.Y.2d 897, 252 N.E.2d 128, 304 N.Y.S.2d 593 (1969); Ippolito v. Power, 22 N.Y.2d 594, 241 N.E.2d 232, 294 N.Y.S.2d 209 (1968); Nodar v. Power, 18 N.Y.2d 697, 220 N.E.2d 267, 273 N.Y.S.2d 273 (1966), 294 N.Y.S.2d 209, 211.

[39] Ippolito v. Power, 22 N.Y.2d 594, 597, 241 N.E.2d 232, 233 (1968).

[40] M. Finkelstein, *supra* note 3, at 120-28.

judge has accepted the premise that there is no pattern, then the factfinding task is to assess the probability that once the irregular marbles have been withdrawn randomly, the remaining marbles in the urn representing the winner's votes will continue to outnumber the remaining marbles representing the opponent's votes. That probability will clearly vary with the winner's margin of victory and the number of votes that have to be withdrawn.

Intuitive estimates of the probability of the irregular votes changing the election's outcome are subject to the same problem of sample size differences noted previously in the jury discrimination discussion.[41] Finkelstein notes two New York cases which one judge deemed as "almost identical" in facts.[42] In one, however, the number of irregular votes was 109 and the winner's margin of victory was 27 votes.[43] In the other, the number of irregular votes was 136 and the winner's margin was 62 votes.[44] The probability of the irregular votes determining the election in the first case was calculated by Finkelstein to be approximately one in a hundred. In the second situation, he found the probability to be less than one in a million. Whether this difference would have been persuasive to the judicial factfinder is unknown because apparently such information was not made available to the court. The greater precision offered by Finkelstein's recommended calculation[45] would undoubtedly be useful for the judge to see[46] since the facts of the two cases were not "almost identical" from a statistical point of view.

III. Dangers in Innovative Models for the Intuitive Factfinder

The focus of the preceding section of this review has been on the value of providing a precise method of calculating probabilities when the express legal question is phrased in straightforward terms of probability. Such responsive uses of quantitative methods, such as in the jury discrimination cases, provide more accurate and consistent results. There are dangers inherent in more innovative uses of statistical methods in the

---

[41] *See* text accompanying notes 23-27 *supra.*

[42] M. Finkelstein, *supra* note 3, at 120-28.

[43] Ippolito v. Power, 22 N.Y.2d 594, 596, 241 N.E.2d 232, 233, 294 N.Y.S.2d 209, 211 (1968).

[44] DeMartini v. Power, 27 N.Y.2d 149, 151, 262 N.E.2d 857, 857, 314 N.Y.S.2d 609, 610 (1970).

[45] Finkelstein's method of analysis, like the one he employs in jury discrimination cases, is based on sampling: what is the probability that the true composition of the irregular votes—a sample from all the possible combinations of votes—would be as extreme as necessary to cause a reversal in the election? If the probability is below the critical value, such as 0.05, the court may choose to characterize the chance as too rare. The calculation is different from the one in Finkelstein's jury chapter, because it is based on a hypergeometric distribution, appropriately accounting for the fact that the sampling is without replacement from a finite pool. This adjustment is necessary particularly when the number of votes is small, but when large numbers are involved the binomial distribution, as used for the jury discrimination analysis, could be used without serious error. *See* H. Blalock, *supra* note 30, at 170-72.

[46] Finkelstein notes that in DeMartini v. Power, 22 N.Y.2d 594, 241 N.E.2d 232 314 N.Y.S.2d 609 (1968), the New York Court of Appeals had the assistance of statistical analysis in the appellant's brief when it reversed the Appellate Division in that case. M. Finkelstein, *supra* note 3, at 124 & n.35.

courtroom, however. One danger is that an innovation in finding an answer can have the effect of altering the legal question.[47] Another danger is that the introduction of probability to correct intuitive errors in only one part of a fact-finding process may actually compound the effect of other uncorrected intuitive errors.[48] An illustration of this process is in the following discussion of Finkelstein's suggested use of Bayes' theorem for the integration of forensic evidence with other evidence in criminal trials.

Finkelstein was inspired for this chapter on criminal evidence by the work of Kahneman and Tversky.[49] These researchers have found, in another series of studies, that individuals tend to ignore background information on relative probabilities and rely instead on information relating directly to stereotyped notions.[50] In one study, for example, Kahneman and Tversky asked subjects to indicate the likelihood that a particularly described person was an engineer or a lawyer. Subjects were told that a group of seventy lawyers and thirty engineers were in a group and that thumbnail descriptions were made on each of the one hundred. The description of one person included information that he was married with four children, conservative, uninterested in politics and enjoyed home carpentry as a hobby. Because this information fits with a stereotyped image of an engineer more than a lawyer, the subjects tended to ignore the fact that far more lawyers than engineers were included in the original group. This reliance on stereotypes persisted even when the source of the information used was generally considered worthless.[51]

Noting such findings, Finkelstein wrote an article in 1970,[52] now chapter three in *Quantitative Methods in Law*, advocating the use of Bayes' theorem to assist jurors in criminal trials when they must integrate prior evidence of guilt with some other probabilistic evidence, such as the probability that the defendant's fingerprint was found on the murder weapon. The hypothetical case he used involved a defendant charged with murdering his girlfriend. The evidence included a violent quarrel between the deceased and the defendant the night before, a history of the defendant striking the deceased and a partial fingerprint. Expert testimony on the fingerprint indicates that the partial print matches that of the defendant, and such prints appear in no more than one case in one thousand. The technique Finkelstein advocates to integrate this quantitative evidence with the nonquantitative testimony is use of Bayes' theorem. Jurors would determine their degree of belief in the defendant's guilt without the fingerprint evidence, such as one in four, and then with the

[47] *See* Tribe, *supra* note 5. *See also* the discussion of models of jury representativeness versus jury discrimination at notes 31-36 and accompanying text *supra*.

[48] *See* text accompanying notes 63-64 *infra*.

[49] M. Finkelstein, *supra* note 3, at 71-72.

[50] Kahneman & Tversky, On the Psychology of Prediction, 80 Psychological Rev. 237, 239 (1973).

[51] *Id.* at 241-43.

[52] Finkelstein, *supra* note 5.

assistance of expert testimony would be told how that prior probability combines with the one in a thousand fingerprint statistic. If the prior assessment was one in four, for example, the combined probability could be seen to be 997 out of 1000. As Finkelstein notes, this high combined probability is likely to be surprising to jurors because of the natural tendency to underestimate combinatorial effects.[53]

Aside from any objection that this process may have changed the nature of the trial,[54] there is a hidden danger in this approach that might lead jurors to overestimate the base rate probabilities on which Finkelstein's use of Bayes' theorem relies. The process of overestimating the base rate, also found in the work by Kahneman and Tversky, is explained in the next section. Then, the final section in this review explains why these conflicting intuitive biases produce an undesirable result in Finkelstein's hypothetical trial.

A. *Errors in Base Rate Estimation*

One kind of probabilistic task performed by legal decisionmakers involves the estimation of base rates of probability. This task is in contrast to the previously discussed problems of assessing deviations from established base rate expected values. Situations involving tossing coins, the birth of baby boys versus girls, or the distribution of minorities on juries all have established base rates, such as half of coin flips are expected to be heads, from which a certain degree of deviation occurs. Often there is no base rate already established from which one can ascertain rareness, so the base rate must be estimated. For example, how likely is it that a car will roll downhill unless the driver failed to set the handbrake? Individuals must assess frequency from experience, which involves thinking of instances of the event and then determining frequency from how easily instances come to mind. When asked to estimate the divorce rate, for example, one thinks of how many acquaintances have recently obtained divorces and then makes a large or small estimate accordingly.[55]

Researchers Kahneman and Tversky have found that this estimation process is subject to serious bias. It was previously noted that they identified one source of bias in making subjective probability assessments as an excessive natural reliance on the supposed representativeness of certain configurations. A second source of bias emerges in estimating probability from frequencies when certain events are more easily recalled than others. Airplane crashes are such noteworthy events that they tend to remain salient in individuals' minds. Thus, the probability of airplane crashes tends to be overestimated. Similarly, the likelihood of a nuclear accident would be subjectively overestimated immediately following such an acci-

---

[53] M. Finkelstein, *supra* note 3, at 92-93 n.57. *See also,* Kaplan, Decision Theory and the Factfinding Process, 20 Stan. L. Rev. 1065, 1084-86 (1968).

[54] *See* the Finkelstein-Tribe debate, *supra* note 5.

[55] Tversky, *supra* note 1, at 208.

dent, or following a popular movie making such an event a more salient possibility in people's minds.[56]

One study that illustrates this type of bias derived from the use of this availability heuristic involved asking people whether there are more words beginning with the letter "k" or having "k" in the third letter position. Individuals consistently responded that there are more words beginning with "k," such as kick, than having "k" in the third position, such as like. A systematic count, however, reveals that "k" appears as the third letter about twice as often as the first letter in a word. Individuals were biased in their estimates by the greater ease with which one can think of words starting with certain letters than words with those letters in the middle.[57]

This tendency to judge frequency by the availability of instances that come to mind easily affects other kinds of estimates as well. Individuals asked to estimate quickly a numerical expression such as $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$ judge the answer larger than the expression $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$. The researchers found that in their study the average guess for the first expression, with the numbers in descending order, was 2250. The average guess for the numbers in ascending sequence was 512. Both answers were gross underestimates of the actual calculated answer, which is 40,320. The researchers explain this difference between the descending and ascending orders as a bias caused by extrapolating from the first available numbers.[58] The results of the first few steps in the multiplication for the descending order suggest to the estimator that the final result is going to be relatively large, whereas the first few steps in the ascending order—1×2 is 2, then 2×3 is 6—suggest incorrectly that the overall expression is going to be small.[59]

This selectivity bias occurs even among those trained as professionals in the subject matter in question. In another study Kahneman and Tversky found that clinical psychologists would overestimate the likelihood of suicide by a depressed patient. The overwhelming majority of depressed patients never commit suicide, but suicide is a dramatic and memorable event and is thus easily called to mind when estimating the probability of suicide. The researchers speculate that the same process is likely to bias the intuitive predictions of other professionals as well, such as stockbrokers, sportscasters, political analysts, or research psychologists.[60] Lawyers,

---

[56] *Id.* at 230.

[57] *Id.* at 211-12.

[58] *Id.* at 215-16.

[59] Finkelstein argues for greater precision in legal fact-finding tasks involving related kinds of numerical estimates in his chapter on administrative rate making. Using an example of administrative rate setting, he maintains that if the agency decisionmakers wish to consider such factors as comparative differences in risks among regulated utilities, intuition provides a poor guide for arriving at well-informed decisions. The chapter details the possible use of regression analysis in such cases and notes that unlike intuitive methods, no valuable information from the data needs to be discarded in order to reduce the numbers to manageable terms. M. Finkelstein, *supra* note 3, chapter 7 ("Regression Models in Administrative Proceedings") *originally appearing* in 86 Harv. L. Rev. 1442 (1973).

[60] Tversky, *supra* note 1, at 227-30.

judges and juries are prone to the same fallacies too; professionalism and good intentions do not appear to be sufficient to overcome these natural biases.

B. *Compounding Biases: A Solution Providing No Cure*

The work of Kahneman and Tversky described in this review has identified two devices relied upon by individuals to obtain estimates of likelihood. The availability heuristic was shown to influence judgments about the frequency of events in that an event is deemed more frequent, and thus more probable, if instances are easily recalled.[61] The other heuristic, representativeness, leads individuals to judge events as more probable if they comport with stereotyped ideas about commonness.[62] Judgments based on these biases are sometimes reasonable, but often they result in severe and systematic errors in estimation.

One feature of individuals' reliance on the representativeness heuristic that Finkelstein noted in particular is the tendency to ignore background information on relative probabilities and to rely instead on information relating directly to stereotyped notions.[63] This problem inspired Finkelstein to propose the use of Bayes' theorem in criminal trials to assist jurors in integrating evidence. In his example jurors would assess the prior probability of the defendant's guilt based on testimony concerning matters such as the previous fight between the victim and the defendant. Then Bayes' theorem is used to integrate that assessment with the probabilistic fingerprint evidence.

It is true that the research Finkelstein cites indicates that individuals are poor at integrating prior probabilities with new evidence. It is also likely that jurors would underestimate the combined probabilities. Finkelstein's solution, however, requires jurors to perform a task in which they are likely to *over*estimate the defendant's guilt if the evidence comports with sterotyped notions concerning murders or murderers. The Finkelstein approach requires jurors to estimate the prior probability of guilt based on the nonquantitative evidence about the fight the night before and the history of striking the girlfriend. The bias created by the availability heuristic would tend to make jurors overestimate the defendant's guilt based on the evidence. Even though most boyfriends who quarrel or strike their girlfriends do not then murder them, instances where such events do lead to murder are likely to be more salient in the jurors' minds. More salient events are estimated as more frequent ones, so the probative value of this evidence is likely to be overestimated.[64]

The effect of overestimating the probability of prior guilt counterbalances to some degree the effect of underestimating the combinatorial effect of the fingerprint evidence. Although there is no way of being sure

---

[61] *See* text accompanying notes 55-60 *supra*.
[62] *See* text accompanying notes 25-29 *supra*.
[63] *See* notes 49-51 and accompanying text *supra*.
[64] *See* notes 55-60 and accompanying text *supra*.

that the resulting verdict from these combined intuitive processes is statistically defensible, it seems unwise in any event to tamper with only one part of this double estimation process. In a case such as the hypothetical one dealt with by Finkelstein, it would appear that, all other considerations aside, half a cure for the fallacies of the human estimation biases is worse than no cure at all.

IV. Conclusion

Michael Finkelstein's book *Quantitative Methods in Law* should be recognized as making a significant contribution to an overly neglected area of the law. His responsive and innovative uses of statistics to address legal questions deserve close attention by all those considering fresh approaches to old as well as new problems. This review has focused primarily on the responsive uses of probability identified in various parts of Finkelstein's work. These uses have been contrasted with the results obtained when the factfinder relies on intuitive estimates of probability. Biases emerging from the natural process of estimation have been identified and this review has argued that these biases should not be allowed to dictate results when the appropriate legal question is a probabilistic one. On the other hand, it is urged that statistics should not be introduced to correct the biases of a criminal jury in a case such as the one hypothesized in the Finkelstein-Tribe debate because the correction process itself fails to eliminate a countervailing bias against the defendant.

Finkelstein's work accentuates some important issues in the legal factfinding process. Although some of his innovative uses of statistics are highly controversial, the issues he raises are worthy of careful consideration by the legal profession. Particularly noteworthy are his suggestions on how to approach with greater precision those legal questions that are essentially probabilistic ones. Subjective estimations are prone to serious errors and are a poor substitute for more exact calculation.