

2002

Probability and Statistics in the Legal Curriculum: A Case Study in Disciplinary Aspects of Interdisciplinary

Mike Townsend

Thomas Richardson

Follow this and additional works at: <https://dsc.duq.edu/dlr>



Part of the [Law Commons](#)

Recommended Citation

Mike Townsend & Thomas Richardson, *Probability and Statistics in the Legal Curriculum: A Case Study in Disciplinary Aspects of Interdisciplinary*, 40 Duq. L. Rev. 447 (2002).

Available at: <https://dsc.duq.edu/dlr/vol40/iss3/3>

This Article is brought to you for free and open access by Duquesne Scholarship Collection. It has been accepted for inclusion in Duquesne Law Review by an authorized editor of Duquesne Scholarship Collection.

Probability and Statistics in the Legal Curriculum: A Case Study in Disciplinary Aspects of Interdisciplinarity

Mike Townsend¹ and Thomas Richardson²

I. INTRODUCTION

I do think . . . that the law was too parochial twenty-five years ago and that despite all the false starts and silly fads that have marred its reaching out to other fields, the growth of interdisciplinary analysis has been a good thing, which ought to (and will) continue. Disinterested legal-doctrinal analysis of the traditional kind remains the indispensable core of legal thought Nevertheless, it seems unlikely that we shall soon (if ever) return to a serene belief in the law's autonomy³

It is easy enough to grant Posner's general assertion about interdisciplinarity, but it is much more difficult to describe meaningful classroom implementations. This Article provides a general taxonomy for various types of educational interdisciplinarity. The discussion focuses, however, on what is called crossdisciplinary education, using a particular mathematics-based example to illustrate how crossdisciplinary education can be used to ask law students to (re)examine law as a discipline.⁴ This is the third installment of a work in progress, the goal of which is not so much to construct a definitive portrait of

1. Associate Professor of Law, University of Washington; B.A. 1973, University of Michigan; M.A. 1978, University of Michigan; Ph.D. 1982, University of Michigan; J.D. 1989, Yale University.

2. Associate Professor of Statistics, University of Washington; B.A. 1992, Oxford; M.Sc. 1995, Carnegie Mellon University; Ph.D. 1996, Carnegie Mellon University.

3. Richard A. Posner, *The Decline of Law as an Autonomous Discipline: 1962-1987*, 100 HARV. L. REV. 761, 777-78 (1987).

4. This Article treats statistics as a part of mathematics, although many consider statistics to be a separate discipline that is part of the "mathematical sciences." In addition, there is no real distinction made here between probability and statistics, although, for example, some see probability as essentially deductive in its application, and statistics as essentially inductive. See, e.g., VIC BARNETT, *COMPARATIVE STATISTICAL INFERENCE* 8 (3d ed. 1999).

law, as it is to examine what such a construction entails.⁵

A consideration of interdisciplinarity must begin with a discussion of the idea of a discipline. Section A of Part II describes a "component-based" notion of discipline, the premise of which is that the basic goal underlying the Western intellectual tradition is to understand, appreciate, and utilize our environment. Understanding (i.e. science) involves classification, appreciation (i.e. art) involves interpretation, and utilization (i.e. technology) involves the means for providing sustenance and comfort. The environment, however, is both complex and textured, and thus the intellectual tradition centers on disciplines, which provide a more focused approach to the basic goal. A discipline is at once a science, an art, and a technology. The science component of a discipline is characterized by the objects considered, the properties studied, and the classification employed. The range of interpretations given, and the symbolic medium used, characterize the art component of a discipline. Finally, the technology component is characterized by the methods and scope of its applications. Within each discipline, the scientific, artistic, and technological components work together to present a specific (part of a) worldview.

Interdisciplinarity, the use of more than one discipline, is important because it sharpens the resolution of the intellectual picture by applying the perspectives of differing disciplines to the most interesting aspects of the environment. Section B of Part II presents a taxonomy of educational interdisciplinarity. In *multidisciplinary education*, the student is exposed to a number of disciplines. However, there is no essential integration in that their relation, if any, does not reach a point where the student feels that studying one discipline presupposes knowledge of another. In *pluridisciplinary education*, the student does recognize that studying one discipline presupposes some (possibly non-trivial) acquaintance with another, but one discipline remains dominant. In *crossdisciplinary education*, no discipline is dominant, but integration does not generalize past a specific setting. Generalization does occur at the fourth, and highest, level of interdisciplinarity, but it can take two forms. In *interdisciplinary*

5. For previous installments, see Mike Townsend, *Cardozo's Allegheny College Opinion: A Case Study in Law as an Art*, 33 HOUS. L. REV. 1103 (1996)[hereinafter Townsend, *Cardozo*] and Mike Townsend, *Implications of Foundational Crises in Mathematics: A Case Study in Interdisciplinary Legal Research*, 71 WASH. L. REV. 51 (1996)[hereinafter Townsend, *Mathematics*].

education, the generalization proceeds so that a new discipline begins to emerge at the intersection of several others. In this Article, the word "*interdisciplinarity*" refers to the general practice of using more than one discipline, and the word "*interdisciplinary*" refers to a particular way of using more than one discipline. On the other hand, *transdisciplinary education* seeks to move beyond boundaries altogether by developing an overarching, transcendent framework covering several disciplines.

In terms of the component-based definition presented here, law is a discipline. Earlier articles comment on the art and science components of law. Section C summarizes the relevant discussion and makes some general observations about interdisciplinarity in legal education.

This Article considers interdisciplinarity and the legal curriculum in the context of probability and statistics. Section D of Part II begins the discussion by sketching some multidisciplinary, pluridisciplinary, interdisciplinary, and transdisciplinary approaches.

Part III is the workhorse of this Article. The particular example used here is the well-known jury discrimination case of *Castaneda v. Partida* as described in Section A. This "case study" provides the basis for a crossdisciplinary experience that offers students an opportunity to think about law as a discipline. It is difficult for students to step back and look at law as a discipline when there is no "back." The idea of the type of crossdisciplinary education described here is not to make law students intelligent consumers of another discipline, but to provide students with another vantage point for thinking about law. That is, the perspectives given by another discipline can be used to reinforce law as a discipline. This is what is meant by the phrase "disciplinary aspects of interdisciplinarity" appearing in the title. Studying the various mathematical techniques described in Sections B and C provides the vantage point here. As will be seen in Section D, students can be asked to think about the individual disciplinary components of law and their relationship, as well as the evolution of law as a discipline. Moreover, students can be asked to think about specific legal doctrine such as the nature of the prima facie case, legal reasoning, and the presumption of innocence in the disciplinary framework.

This Article concludes with some brief observations on the importance of thinking about law as a discipline.

II. THE LAW: DISCIPLINARITY AND INTERDISCIPLINARITY

Man is a singular creature. He has a set of gifts which make him unique among the animals: so that, unlike them, he is not a figure in the landscape—he is a shaper of the landscape. In body and in mind he is the explorer of nature

. . . Man is distinguished from other animals by his imaginative gifts . . . [so that] the great discoveries of different ages and different cultures, in technique, in science, in the arts, express in their progression a richer and more intricate conjunction of human faculties, an ascending trellis of his gifts.⁶

A. *Disciplines*

The premise for the component-based notion of discipline used here is that the basic goal underlying the Western intellectual tradition is to understand, appreciate, and utilize our environment.⁷

Understanding refers to *science*. This use of the word “science” connotes systemization and organization, as opposed to a more narrow association with what usually are called the natural sciences.⁸ Science involves the classification of the objects appearing in the environment according to their important properties.⁹ These objects may be sensory or non-sensory, and the exact nature of the classification, such as description, prediction, prescription, or explanation, depends on the context.¹⁰ This admittedly is an older use of the word “science,”¹¹ but it captures

6. Jacob Bronowski, *THE ASCENT OF MAN* 19-20 (1973).

7. Nothing said here is meant to imply that other traditions lack these concepts.

8. For a discussion of this distinction, see David A. Funk, *Juridical Science Paradigms as Newer Rhetorics in 21st Century Jurisprudence*, 12 N. KY. L. REV. 419, 435 (1985).

9. See WEBSTER'S NINTH NEW COLLEGIATE DICTIONARY 1051 (1987) (giving as one definition “a department of systematized knowledge as an object of study <the [science] of theology>”). This Article does not address debates about the ontology of the objects “appearing” in the environment.

10. An introduction to a general notion of classification is provided in Stephan Körner, *Classification Theory*, in 4 ENCYCLOPEDIA BRITANNICA 691 (15th ed. 1974). For more detailed discussions in specific contexts, see RICHARD B. BRAITHWAITE, *SCIENTIFIC EXPLANATION: A STUDY OF THE FUNCTION OF THEORY, PROBABILITY AND LAW IN SCIENCE* 9-10 (1953) (discussing scientific laws); CARL G. HEMPEL, *ASPECTS OF SCIENTIFIC EXPLANATION AND OTHER ESSAYS IN THE PHILOSOPHY OF SCIENCE* 135-73 (1965) (discussing natural and social sciences); W.S. JEVONS, *THE PRINCIPLES OF SCIENCE* 673-734 (2d ed. 1877) (discussing natural sciences); THOMAS MUNRO, *THE ARTS AND THEIR INTERRELATIONS* (2d ed. 1967) (discussing arts).

11. Some even view this use as “obsolete.” See, e.g., Barbara J. Shapiro, *Law and Science in Seventeenth-Century England*, 21 STAN. L. REV. 727, 727 (1969).

the essence of one of the three basic dimensions of the intellectual tradition.

Appreciation refers to *art*. This use of the word “art” connotes aesthetics, as opposed to a more narrow association with what usually are called the fine arts. Art involves the interpretation of the environment through the creative use of a symbolic medium.¹² As with the concept of science, there is no attempt here to describe these ideas completely. In particular, no effort is made to develop the concepts of the artist, the work of art, and the spectator.¹³ Nonetheless, as with science and technology, the intent is to present notions that cut across disciplines.¹⁴

Finally, utilization refers to *technology*. This use of the word “technology” is intended to connote application, as opposed to a more narrow association with what usually are called the engineering sciences. Technology involves the means employed to provide sustenance and comfort.¹⁵

The environment is both complex and textured, and thus the intellectual tradition centers on disciplines, which provide a more focused approach to the basic goal described above. A discipline is at once a science, an art, and a technology.¹⁶ A discipline is characterized as a science by the objects considered, the properties studied, and the classification employed. As an art, the range of interpretations and the symbolic medium used characterize a discipline. Finally, a discipline is characterized as a technology by the methods and scope of its applications. Within each discipline,

12. See J. FISHER, REFLECTING ON ART 324 (1993) (stating “[f]our major concepts dominate traditional aesthetics: beauty, representation, expression, and form”); Dewitt H. Parker, *Aesthetics*, in ENCYCLOPEDIA OF THE ARTS 14, 15 (Dagobert D. Runes & Harry G. Schrickel eds., 1946) (stating “[t]hree general conceptions of art have dominated the history of aesthetics: *imitation, imagination, and expression or language*”).

13. Any such effort would move far beyond the scope of this Article and into controversial areas. See HUGH CURTLER, WHAT IS ART? 1, 3 (Hugh Curtler ed., 1983).

14. Cf. HELEN GARDNER, UNDERSTANDING THE ARTS 318-28 (1932). No attempt is made here to define the term “art” as it is used in connection with music, painting, etc. Indeed, there is perhaps nothing more problematic. See HORST W. JANSON, HISTORY OF ART 9 (1967). For a discussion of the problems involved, see Paul Ziff, *The Task of Defining a Work of Art*, 62 PHIL. REV. 58 (1953). For a collection of classic perspectives, see FRANK A. TILLMAN & STEVEN M. CAHN, PHILOSOPHY OF ART AND AESTHETICS FROM PLATO TO WITTGENSTEIN (1969). For general overviews, see MONROE C. BEARDSLEY, AESTHETICS FROM CLASSICAL GREECE TO THE PRESENT: A SHORT HISTORY (1966); MUNRO, *supra* note , at 49-109.

15. See WEBSTER'S NINTH NEW COLLEGIATE DICTIONARY 1211 (1987) (defining technology as “the totality of the means employed to provide . . . human sustenance and comfort”).

16. For another component-based approach to the notion of a discipline, see Timothy P. Terrell, *Flatlaw: An Essay on the Dimensions of Legal Reasoning and the Development of Fundamental Normative Principles*, 72 CAL. L. REV. 288 (1984).

the scientific, artistic, and technological components work together¹⁷ to present a specific (part of a) worldview.¹⁸

B. *Interdisciplinarity*

Interdisciplinarity, using more than one discipline, sharpens the resolution of the intellectual picture of the environment by applying different perspectives to the most interesting aspects of the environment. The major intellectual challenge for interdisciplinarity today is fostering meaningful work in an information-rich era in which it is difficult to master even a small part of a given field.

While the idea of interdisciplinarity can be traced back to Plato, it was not until the latter part of this century that interdisciplinarity was studied *per se*.¹⁹ Although frameworks for interdisciplinarity vary,²⁰ the hierarchy presented here emphasizes the level of integration rather than its operation or purpose.²¹ Moreover, this hierarchy is described in educational, rather than research, terms.

In *multidisciplinary education*, the student is exposed to a number of disciplines, but there is no essential integration in that their relation, if any, does not reach a point where the student feels that studying one discipline presupposes knowledge of another. For this reason, multidisciplinary education often is left largely to the student. The obvious example is the stand-alone undergraduate distribution requirement.

In *pluridisciplinary education*, the student does recognize that

17. Others posit an antagonistic relationship. Cf. DANIEL BELL, *THE COMING OF POST-INDUSTRIAL SOCIETY* 374-77 (1973) (discussing potential antagonism between "scientific, technological, and cultural estates"); THE REPUBLIC OF PLATO b.X (Francis M. Cornsford trans., 1941) (discussing potential corrupting influence of fine arts on search for knowledge).

18. What does it mean, after all, to say that students are taught to think like a lawyer? The notion of the separation of disciplines can be traced back at least as far as Aristotle. See T.Z. LAVINE, *FROM SOCRATES TO SARTRE: THE PHILOSOPHIC QUEST* 76 (1984). The current scope and organization of disciplines has been affected in part by the social and institutional factors that accompanied the transition from the educated amateur, to the professional society, to the modern research university. See ROGER L. GEIGER, *TO ADVANCE KNOWLEDGE: THE GROWTH OF AMERICAN RESEARCH UNIVERSITIES 1900-1940* 20-27 (1986). For discussions of the American version of this transition, see *THE ORGANIZATION OF KNOWLEDGE IN MODERN AMERICA 1860-1920* (Alexandra Oleson & John Voss eds., 1979); *THE PURSUIT OF KNOWLEDGE IN THE EARLY AMERICAN REPUBLIC: AMERICAN SCIENTIFIC AND LEARNED SOCIETIES FROM COLONIAL TIMES TO THE CIVIL WAR* (Alexandra Oleson & Sanborn C. Brown eds., 1976).

19. For a discussion of the evolution of interdisciplinarity, see JULIE THOMPSON KLEIN, *INTERDISCIPLINARITY: HISTORY, THEORY, AND PRACTICE* 19-39 (1990).

20. See KLEIN, *supra* note 19, at 55.

21. The terminology is based loosely on Joseph J. Kockelmans, *Why Interdisciplinarity?*, in *INTERDISCIPLINARITY AND HIGHER EDUCATION* 123, 127-29 (Joseph J. Kockelmans ed. 1979) [hereinafter KOCKELMANS, HIGHER], and, to a lesser extent, KLEIN, *supra* note 19, at 55-73.

studying one discipline presupposes some (possibly non-trivial) acquaintance with another discipline, but one discipline remains dominant. The typical example involves a service discipline such as mathematics used by a client discipline such as physics. At its core, pluridisciplinary education forces students to realize that they must be “intelligent consumers” of other disciplines.

In *crossdisciplinary education*, no discipline is dominant, but the integration does not generalize past a specific setting. For example, a student writing a paper on housing problems might find it necessary to integrate materials from a number of disciplines, yet the integration does not extend beyond the specific context.

At the fourth, and highest, level of interdisciplinarity, general integration does occur, but it can travel in one of two directions. In *interdisciplinary education*, the generalization proceeds in such a way that a new discipline begins to emerge at the intersection of others. A biochemistry Ph.D. program is an example of the mature stage of this process. On the other hand, *transdisciplinary education* seeks to de-emphasize existing disciplinary boundaries by developing an overarching, transcendent framework covering a number of disciplines.²² There are, for example, “schools of information science” that treat information creation and management as unifying concepts cutting across disciplines such as computer science and library science.²³

C. *The Law*

In terms of the component-based notion of discipline presented here, to say that law is a discipline is to assert that law is a science, an art, and a technology. The resulting legal worldview is not only what makes it possible to write a book on comparative law, but also what makes it easier for a lawyer than a mathematician to digest it. Consider each assertion in turn.

The core of understanding (i.e. science) is classification. There is no attempt here to provide any classification scheme for law. Law students are well aware of the sort of pigeonholing typifying their doctrinal and clinical courses. The assertion is merely that classification is the heart of the scientific component of law as a

22. The notions of art, science, and technology are not transdisciplinary, but rather parts of the definition of discipline itself. An earlier article critiques a particular example of transdisciplinary research. See Townsend, *Mathematics*, *supra* note 5.

23. See Katherine S. Mangan, *In Revamped Library Schools, Information Trumps Books*, *Chronicle of Higher Education*, THE CHRONICLE OF HIGHER EDUCATION, April 7, 2000, at A43.

discipline.²⁴ The idea that law involves classification is nothing new, and classification is an important part of both the common and civil law traditions.²⁵ This idea of law as a science is explored more in a previous article.²⁶

As an art, law involves the creative use of an interpretive medium. An earlier article discusses the idea of law as an art in the context of contract law.²⁷ The article argues that theories of promissory liability are interpretations in the legal media of (some of the) typical answers given by students to the question of why particular promises should be kept. For example, the bargain theory is the legal interpretation of the answer "We had a deal!" Cardozo's *Allegheny College* opinion creatively juxtaposes these theories with an effect reminiscent of that produced by parts of Mark Antony's funeral oration for Julius Caesar, his purpose being to buttress the nascent Restatement support for reliance as an independent basis of promissory liability.

The idea that law is a technology is perhaps the least controversial of the assertions and will not be discussed here in any detail.

Posner is correct in saying that modern legal education will continue to draw on a wide range of disciplines, yet one might go further. Some scholars assert that certain disciplines, such as philosophy, involve more interdisciplinarity than others.²⁸ Law is another such discipline. Indeed, one can argue that the development of American legal education since 1870 is in fact the history of three waves of interdisciplinarity; one based on the natural sciences and mathematics as typified by Langdell's Harvard; one based on the social sciences as typified by the Legal Realists; and one on the humanities as typified by the current interest in "texts" and "narratives".

One can find all levels of the interdisciplinarity hierarchy in the modern law school. For many students, legal education is

24. For an exhaustive and provocative account of classification in law, see Jay M. Feinman, *The Jurisprudence of Classification*, 41 STAN. L. REV. 661 (1989). For a general discussion of the older conception of the word "science" in connection with law, see Harold J. Berman & Charles J. Reid, Jr., *The Transformation of English Legal Science: From Hale to Blackstone*, 45 EMORY L.J. 437 (1996).

25. For an overview of various schemes of legal classification, see 5 ROSCOE POUND, JURISPRUDENCE 5-75 (1959).

26. See Townsend, *Mathematics*, *supra* note 5.

27. See Townsend, *Cardozo*, *supra* note 5.

28. See Carl R. Hausman, *Disciplinarity or Interdisciplinarity?*, in KOCKELMANS, HIGHER, *supra* note 21, at 1, 8-9.

multidisciplinary in that they can apply non-law courses toward their degree. In many instances, law and economics is treated in a pluridisciplinary manner. For example, antitrust casebooks usually begin with some discussion of microeconomics. Many topic-oriented courses and seminars are crossdisciplinary. A seminar in medical ethics, for example, might integrate law, philosophy, medicine, and public health in a crossdisciplinary manner.²⁹ Jurisprudence can be treated in an interdisciplinary manner.³⁰ Finally, the notion of text and narrative is in vogue as a transdisciplinary concept unifying a wide range of disciplines including law.³¹

D. Probability and Statistics in the Legal Curriculum

“For the rational study of the law the black-letter man may be the man of the present, but the man of the future is the man of statistics and the master of economics.”³²

Probability and statistics can be integrated into the legal curriculum at all levels of the interdisciplinarity hierarchy.³³ Law schools not offering any overt discussion of probability and statistics still can make a multidisciplinary experience possible by counting probability and statistics courses toward the J.D. degree. Law schools can offer students a pluridisciplinary experience. One possibility is a physics/mathematics model in which students take probability and statistics, as part of their prelaw studies, and law schools develop specific probability-and-statistics-based units for inclusion in a wide range of classes such as first-year courses, e.g. torts, traditional electives, and specialized courses, e.g. discrimination law. As an alternative, law schools can teach a required, stand-alone quantitative methods course that presents specific applications from previous law courses.³⁴ To provide an

29. Conversation with Professor Pat Kuszler, August 14, 1997.

30. Indeed, the University of California at Berkeley offers a Ph.D. program in Jurisprudence and Social Policy.

31. See, e.g., PAULINE M. ROSENAU, *POST-MODERNISM AND THE SOCIAL SCIENCES: INSIGHTS, INROADS, AND INTRUSIONS* 25-41, 124-27 (1992). See also, GARY MINDA, *POSTMODERN LEGAL MOVEMENTS: LAW AND JURISPRUDENCE AT CENTURY'S END*, 149-66 (1995) (discussing law and literature movement).

32. Oliver Wendell Holmes, *The Path of the Law*, 10 HARV. L. REV. 457, 469 (1897).

33. For another discussion of the general issue, see David H. Kaye, *Thinking Like A Statistician: The Report of the American Statistical Association Committee on Training in Statistics and Selected Professions*, 34 J. LEGAL EDUC. 97 (1984).

34. Texts that can be used in such a course include DAVID W. BARNES, *STATISTICS AS PROOF: FUNDAMENTALS OF QUANTITATIVE EVIDENCE* (1983); DAVID W. BARNES & JOHN M. CONLEY,

interdisciplinary experience, law schools can offer a course based on the sort of material covered in a traditional forensic science program.³⁵ Law schools also can offer a transdisciplinary seminar based on the so-called new evidence scholarship, which seeks to tie legal notions of evidence to an overarching notion of reasoning encompassing a range of disciplines including law, mathematics, sociology, and philosophy.³⁶

The emphasis here is on crossdisciplinary education. Students consider topics that no particular discipline can adequately address. No discipline is dominant, but the integration of disciplines is not generalized past the context at hand. As with pluridisciplinary education, crossdisciplinary education certainly will suggest to students that they must be intelligent consumers of other disciplines. However, one can pursue a further goal of reinforcing law as a discipline. That is, the perspectives given by other disciplines can be used to reinforce law as an art, a science, and a technology. This is what is meant by the phrase "disciplinary aspects of interdisciplinarity" appearing in the title of this Article.

The remainder of this Article illustrates how one can pursue this goal using the well-known jury discrimination case of *Castaneda v. Partida*.

III. CASTANEDA

"The degree of underrepresentation must be proved . . . by comparing the proportion of the group in the total population to the proportion called to serve as grand jurors over a significant period of time."³⁷

STATISTICAL EVIDENCE IN LITIGATION (1986); STEVEN M. CRAFTON AND MARGARET F. BRING, QUANTITATIVE METHODS FOR LAWYERS (1994); WAYNE C. CURTIS, STATISTICAL CONCEPTS FOR ATTORNEYS: A REFERENCE GUIDE (1983); MICHAEL O. FINKELSTEIN & BRUCE LEVIN, STATISTICS FOR LAWYERS (1990); PHILLIP I. GOOD, APPLYING STATISTICS IN THE COURTROOM: A NEW APPROACH FOR ATTORNEYS AND EXPERT WITNESSES (2001); STATISTICS AND THE LAW (Morris H. DeGroot, Stephen E. Fienberg & Joseph B. Kadane eds. 1986).

35. Texts that can be used in such a course include C.G.G. AITKEN, STATISTICS AND THE EVALUATION OF EVIDENCE FOR FORENSIC SCIENTISTS (1995); DAVID H. KAYE, SCIENCE IN EVIDENCE (1997); FEDERAL JUDICIAL CENTER, REFERENCE MANUAL ON SCIENTIFIC EVIDENCE (1994); ANDREA A. MOENSSENS, JAMES E. STARRS, CAROL E. HENDERSON & FRED E. INBAU, SCIENTIFIC EVIDENCE IN CIVIL AND CRIMINAL CASES (4th ed. 1994); BERNARD ROBERTSON & G.A. VIGNAUX, INTERPRETING EVIDENCE: EVALUATING FORENSIC SCIENCE IN THE COURTROOM (1995).

36. For an introduction to this scholarship, see Richard Lempert, *The New Evidence Scholarship: Analyzing the Process of Proof*, 66 B.U.L. REV. 439 (1986). Texts that can be used in such a course include DAVID A. SCHUM, EVIDENTIAL FOUNDATIONS OF PROBABILISTIC REASONING (1994); JONATHAN L. COHEN, THE PROBABLE AND THE PROVABLE (1977).

37. *Castaneda v. Partida*, 430 U.S. 482, 494 (1977).

A. *Background*

Discrimination is a good topic for crossdisciplinary study. There are several relevant disciplines (e.g. law, mathematics, history, political science), and it is difficult to label any one discipline dominant. The focus here is on two of the relevant disciplines, law and mathematics.

The specific example chosen for crossdisciplinary study is the *Castaneda* case. For the purposes of this Article, the *Castaneda* case can be described as follows.³⁸ Plaintiff alleged a violation of his Constitutional rights in that his indictment was brought under a grand jury system systematically excluding Mexican-Americans.³⁹ According to the Supreme Court, "substantial underrepresentation . . . constitutes [a 14th Amendment equal protection violation] if it results from purposeful discrimination."⁴⁰ Moreover, "once [plaintiff] has shown substantial underrepresentation of his group, he has made out a prima facie case of discriminatory purpose, and the burden shifts to the State to rebut that case."⁴¹ To make out the prima facie case,⁴² plaintiff offered in evidence the facts that 79.1% of the county had Spanish surnames, according to a U.S. census, and that only 339 of the last 870 grand jurors chosen were Mexican-American.

The next two subparts describe some of what mathematicians might say about plaintiff's data. In particular, two statistical camps are described: frequentist and Bayesian. Five somewhat parallel evaluative approaches are presented within each of these two camps. The discussion here is not exhaustive, as might well be the case were the primary purpose to produce intelligent consumers of probability and statistics. Rather, the topics are chosen to illustrate how a crossdisciplinary approach can be used to reinforce law as a discipline and encourage students to think about law as a

38. In its broad outlines, this discussion owes much to RICHARD M. ROYALL, *STATISTICAL EVIDENCE: A LIKELIHOOD PARADIGM* (1997), D.H. Kaye, *What is Bayesianism? A Guide for the Perplexed*, 28 *JURIMETRICS* 161 (1988), and David Kaye, *Statistical Evidence of Discrimination*, 77 *J.A.M. STAT. ASSOC.* 773 (1982).

39. 430 U.S. at 490.

40. *Id.* at 493.

41. *Id.* at 494-95. The dissenters questioned whether this was an accurate statement of the law, suggesting that prior case law required an additional showing that the selection process itself provided a clear and easy opportunity for racial discrimination. *Id.* at 510-14. (Powell, J., dissenting).

42. This Article focuses only on what the Court described as the prima facie case. In particular, any discussion of notions such as purpose and causation is beyond the scope of this Article.

discipline.⁴³

B. *The Frequentists*

“[A] majority of authors who present theories of [statistics] favor a frequency interpretation.”⁴⁴

1. *Background*

Some people hold that the “rules” of probability and statistics “apply” only when it is possible to imagine a procedure that can be repeated uniformly an arbitrary number of times. In the disciplinary terms described above, the scientific component of probability and statistics consists of certain definitions, theorems, and proofs (the “probability calculus”), and the technological component depends on repeatability.⁴⁵ According to Richard von Mises:

We state here explicitly: The rational concept of probability, which is the only basis of the probability calculus, applies only to problems in which either the same event repeats itself again and again, or a great number of uniform elements are involved at the same time We will say that a collective is a mass phenomenon or repetitive event, or, simply, a long sequence of observations for which there are sufficient reasons to believe that the relative frequency of the observed attribute would tend to a limit if the observations were indefinitely continued. This limit will be called the probability of the attribute considered within the given collective.⁴⁶

43. Indeed, the emphasis here is not meant to suggest that the frequentist and Bayesian camps exhaust the field of probability and statistics. For example, some assert that the frequentist position on probability has in fact been superseded by the so-called “propensity theory.” See, e.g., DONALD GILLIES, *PHILOSOPHICAL THEORIES OF PROBABILITY* 184 (2000). Nor is the discussion here meant to suggest that probability and statistics exhausts what mathematicians have to say about reasoning in the face of uncertainty. For example, some mathematicians offer chaos theory or fuzzy set theory as an alternative to the probability and statistics approach. See BARNETT, *supra* note 4, at 92-95.

44. Ian Hacking, *Probability*, in *UTILITY AND PROBABILITY* 173 (John Eatwell, Murray Millgate & Peter Newman eds. 1990).

45. Questions about the ontological and epistemological basis of mathematical definitions, theorems, and proofs are beyond the scope of this Article. For a general discussion, see STEPHEN KÖRNER, *THE PHILOSOPHY OF MATHEMATICS AN INTRODUCTION* (1960).

46. Richard Von Mises, *Probability, Statistics, and Truth* 11, 15 (2d ed. 1957). The discussion of probability and statistics here is largely descriptive rather than normative. Moreover, the camps and evaluative approaches are described informally in terms that hide various internal disputes, and the description goes only so far as is necessary for the purposes of this Article. For more complete discussions of frequentist premises, see GILLIES, *supra* note 43, at 88-112; ROY WEATHERFORD, *PHILOSOPHICAL FOUNDATIONS OF PROBABILITY THEORY*

It is important to note the general parameters of such a position. Consider, for example, flipping a coin. A frequentist talks about the probability of heads in terms of a typical, not a particular, observation. That is, the word "probability" refers to the limiting value of the relative frequency of heads that would appear in an unlimited sequence of flips.⁴⁷ For frequentists, however, it is nonsensical to say, "the probability is .5 that heads appeared the first time this coin was flipped." This sentence refers to an isolated nonrepeatable historical event that either did or did not happen. This example indicates why the word "frequentist" is used.

This point of view is termed *objective* in the sense that the concept of "probability" is divorced from consideration of personal beliefs regarding uncertainty (or subjective) factors. Although the frequentist approach is not the only objective approach to probability,⁴⁸ it is chosen here because it is the dominant objective approach encountered in the statistical community.⁴⁹ This perspective is contrasted with a *subjective* approach in Section C below.⁵⁰

Consider how one can abstract the *Castaneda* (under)representation issue so as to make it amenable to a frequentist analysis. The idea is to use the *Castaneda* evidence in conjunction with various *hypotheses* about the representativeness of the selection process. Generally speaking, the hypotheses take the form that the selection process is correctly modeled by randomly selecting balls (with replacement) from an urn containing some proportion of balls labeled MA for Mexican-American.⁵¹ More specifically, the hypothesis is described in terms of an assertion about the proportion of MA balls in the urn.

It is important to understand that a frequentist does not, indeed cannot, make a probability statement about the proportion of MA balls in the urn as there is no repeatability for such a hypothesis per se. Rather, the frequentist must tie nonprobabilistic statements,

144-218 (1982).

47. To emphasize the point made in the last footnote, there is no discussion about the limit process or about the ultimate source of variation in the outcomes of the flips.

48. Another is the so-called propensity approach. For a general discussion, see GILLIES, *supra* note 43, at 113-68.

49. See BARNETT, *supra* note 4, at 70.

50. There is some controversy about whether there are terms more appropriate than "objective" and "subjective." See GILLIES, *supra* note 43, at 18-22.

51. This framework was suggested for the *Castaneda*-type context in Michael O. Finkelstein, *The Application of Statistical Decision Theory to the Jury Discrimination Cases*, 80 HARV. L. REV. 338 (1966).

for example, about the proportion of MA balls, to probabilistic statements about repeatable events. The next few subsections present some standard frequentist approaches along these lines.

2. *Assessing the Strength of the Evidence With Respect to a Single Hypothesis: Fisher's P-Value*

According to the statistician R.A. Fisher, one proper goal when applying probability and statistics is to describe "the strength of the evidence against [a] hypothesis."⁵² There are three things to keep in mind about this approach. Most importantly, this approach is descriptive as opposed to prescriptive. That is, there is no suggestion that one act on the determination made. In this sense, the approach is *inferential*, not *decisionmaking*.⁵³ As an inferential methodology, the p-value approach describes the (strength of the) evidence (relative to a selected hypothesis); it does not assess the validity of the hypothesis. In this sense, the p-value approach may be called *evidential* as opposed to *validational*.⁵⁴ Finally, there is only one hypothesis of interest, no alternative hypothesis is specified.⁵⁵

The hypothesis of interest in the p-value approach is called the *null hypothesis*. As the Fisher quotation indicates, null hypotheses often represent situations with respect to which one wishes to make a negative statement. For example, null hypotheses often describe a sort of *status quo* or "no change" or "nothing new" situation.⁵⁶ With this convention, the fact that 79.1% of the county's population had Spanish surnames, and the Supreme Court's statement of the *prima facie* case, an obvious candidate from the *Castaneda* plaintiff's point of view is the hypothesis that the selection procedure is correctly modeled by randomly selecting balls from an urn (with replacement) with a proportion of MA balls at least .791. This hypothesis is denoted by $\theta \geq .791$.

How does one measure the strength of the evidence relative to

52. R.A. Fisher, *Statistical Methods for Research Workers* 80 (13th ed. 1958).

53. This Article considers the use of probability and statistics for inference and decisionmaking. It does not deal with data summarization. For a brief discussion of data summarization, see LLOYD D. FISHER AND GERALD VAN BELLE, *BIOSTATISTICS: A METHODOLOGY FOR THE HEALTH SCIENCES* 35-74 (1993).

54. For further discussion of the first two distinctions, see Royall, *supra* note 38, at 1-5.

55. For a brief discussion of this point, see BARNETT, *supra* note 4, at 132-33.

56. Mathematical tractability is an additional consideration. Cf. WILLIAM MENDENHALL, *INTRODUCTION TO PROBABILITY AND STATISTICS* 206 (7th ed. 1987) (discussing these considerations in context of hypothesis testing).

this hypothesis? A frequentist approaches this hypothesis by first considering simpler hypotheses of the form $\theta = .791$. For obvious reasons, the latter hypothesis is called *simple*, and a hypothesis of the form $\theta \geq .791$ is called *composite*. Consider then such a simple hypothesis A. The procedure, roughly speaking, is to determine the probabilities of all observations that might have occurred if A were true and report the probability of the collection of observations that are at least as "extreme" as the outcome actually observed. Thus, this procedure involves potential observations that did not occur. The probability is called the *p-value* of the observation and might be symbolized by: P_A (observations at least as extreme as that observed), where the subscript indicates the hypothesis under which the probability is to be evaluated (here the null hypothesis). The p-value is taken to measure the strength of the evidence against the hypothesis on the ground that it indicates the rarity (in the frequency sense) of such an observation.⁵⁷

This strength can be described by comparing the p-value to a qualitative yardstick: a p-value less than .01 is said to indicate very strong evidence against the hypothesis, a p-value greater than or equal to .01 and less than or equal to .05 indicates moderate evidence against the hypothesis, a p-value greater than .05 and less than .1 indicates suggestive evidence against the hypothesis, and a p-value greater than or equal to .1 indicates little or no real evidence against the hypothesis.⁵⁸

It is critical to note that a nonprobabilistic statement about the strength of the evidence (i.e. the outcome actually observed) with respect to a particular hypothesis is tied here to a probabilistic statement about possible outcomes that could occur if the hypothesis were true. This sort of "indirection" typifies the frequentist position, and it should be borne in mind when considering the remaining approaches described in the subsections below.

57. See, e.g., ROYALL, *supra* note 38, at 65. The reader is not alone if she wonders about an evidential approach that uses not only the observation that did occur, but also some that didn't occur. *Id.* at 69.

58. See, e.g., W.J. BURDETTE & E.A. GEHAN, *PLANNING AND ANALYSIS OF CLINICAL STUDIES* 9 (1970). Some suggest dividing these figures in half when employing the one-sided (or "one tail") version of extremity considered here. Cf. Paul Meier, Jerome Sacks & Sandy Zabell, *What Happened in Hazelwood: Statistics, Employment Discrimination, and the 80% Rule*, in *STATISTICS AND THE LAW* 13-15 (M. DeGroot, S. Fienberg & J. Kadane eds. 1994) (discussing issue in context of significance tests). There is a slight twist here as Fisher felt that a p-value too close to 1 was suspect on the ground that data so closely fitting the null hypothesis raised the possibility of tampering with the data. See FISHER & BELLE, *supra* note 53, at 222.

Given *Castaneda's* interest in underrepresentation, it is reasonable to take "extreme" to mean observing no more than 339 Mexican-American grand jurors out of 870 chosen. A standard calculation shows that the probability of seeing no more than 339 MA balls out of 870 chosen from an urn with an MA-proportion equal to .791 is vanishingly small, and thus the *Castaneda* data represents very strong evidence against the hypothesis $\theta = .791$.⁵⁹

What about the composite hypothesis $\theta \geq .791$? In *Castaneda*, it is clear that the p-value with respect to an MA-proportion greater than .791 is less than the p-value with respect to an MA-proportion equal to .791. Since we are assuming that at most one urn correctly models the selection process, one might say that the observation also represents very strong evidence against the composite hypothesis $\theta \geq .791$. In this way, one deals with the composite hypothesis situation in terms of the simple hypotheses contained therein.

3. *Assessing the Strength of the Evidence With Respect to Two Hypotheses: The Law of Likelihood*

Some frequentists deny the validity of the p-value approach. They agree that it is proper to assess the strength of evidence, but they argue that such an assessment can only take place with reference to two hypotheses. Further, they do not accept the use of potential observations that did not occur. Instead of using the p-value approach, they perform the assessment using a certain ratio. According to Richard Royall, this ratio indicates "how observations should be interpreted as evidence for A *vis-à-vis* B, but it makes no mention of how those observations should be interpreted as evidence in relation to A alone."⁶⁰ Note that this approach is inferential, in fact evidential, but it differs from the p-value approach in that there are two hypotheses of interest rather than one. As will be seen, however, this approach does not make any distinction between the two hypotheses of interest by singling one out as any sort of null hypothesis.

What is this ratio, and how does one use it to assess the strength of the evidence relative to two simple hypotheses of the type at issue in *Castaneda*? Given the simple hypothesis A and an

59. Approximately 4.2×10^{-145} . Calculations for this Article were performed with EXCEL. Interestingly, *Castaneda's* approximation is 10^{-140} . See 430 U.S. at 496 n.17.

60. Royall, *supra* note 38, at 8. This approach is characterized here as frequentist, and this seems to reflect the thinking of its main proponents. See *Id.* at xiii-xiv, 169-71; see also A.W.F. EDWARDS, *LIKELIHOOD* xix (expanded ed. 1972).

observation E, the *likelihood* is the probability that such evidence E would have occurred given the hypothesis. Symbolically, the likelihood is $P_A(E)$. Given an observation and two simple hypotheses A and B, the *likelihood ratio* for the hypothesis A versus the hypothesis B is the ratio of the likelihoods. Symbolically, the likelihood ratio is $P_A(E)/P_B(E)$. Roughly speaking, the law of likelihood asserts that the likelihood ratio measures the strength of evidence relative to the two hypotheses.⁶¹ This law is offered on the ground that if some event is more probable under hypothesis A than hypothesis B, then the occurrence of that event is evidence supporting A over B, and the strength of that evidence is measured by how much greater the probability is under A.⁶²

This strength can be described in terms of a qualitative yardstick, but one can also use a "calibrating experiment." Suppose that there are two urns, one containing only MA balls and the other containing an MA-proportion equal to .5. Three balls are chosen from (an unknown) one of the two urns. The balls are MA, MA, MA. The likelihood of this observation with respect to the hypothesis that the balls came from the first urn is 1; the likelihood with respect to the hypothesis that the balls came from the second urn is $1/8$, i.e. $1/2 \times 1/2 \times 1/2$. Therefore, the likelihood ratio for the all-MA urn hypothesis versus the half-MA urn hypothesis is 8. This simple example can be used as a calibration. That is, any likelihood ratio of 8 can be said to have the same evidential import as drawing 3 MAs in a row in this calibration experiment.⁶³ In like fashion, a likelihood ratio of 32 has the same import as drawing 5 MAs in a row, and so on. Using the general result that drawing n MAs in a row yields a likelihood ratio of 2^n , one can even say "a likelihood ratio of 47.5 has the same evidential import as drawing approximately 5.57 MAs in a row in the calibration experiment."

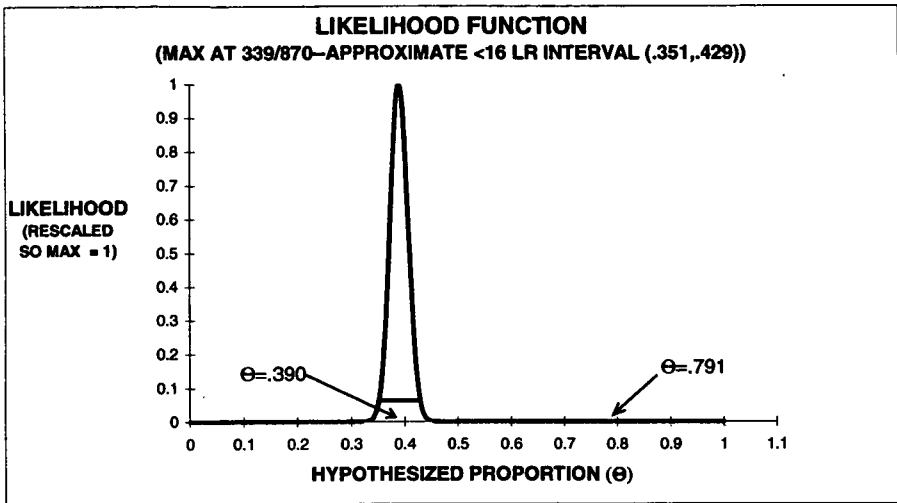
It is natural to compare a given proportion with the proportion that maximizes the likelihood of the observation. In the context of *Castaneda*, one might present the following graph of the likelihoods for an observation of 339 out of 870.

61. See ROYALL, *supra* note 38, at 3. One must distinguish this use of the likelihood ratio from its use in other approaches such as hypothesis testing. See *id.* at 17.

62. *Id.* at 5.

63. *Id.* at 11-12, 24-25.

Figure 1



As should be intuitive, the maximum likelihood occurs at the proportion equal to the observed proportion 339/870 (approximately .390). Any likelihood ratio for two given proportions can be computed by comparing the respective heights of the curve above the horizontal axis. In particular, an observation of 339 jurors out of 870 chosen supports $\theta = 339/870$ (i.e. $\theta = .390$) over $\theta = .791$ to the same extent that an observation of (approximately) 474.7 MAs in a row supports the all-MA urn hypothesis over the half-MA urn hypothesis in the calibration experiment.⁶⁴ As another aid, a horizontal line has been drawn at a height 1/16 that of the maximum height of the curve. Thus, the observation supports the proportion 339/870 over proportions less than or equal to (approximately) .351 or greater than or equal to (approximately) .429 at least to the extent that an observation of 4 MAs in a row supports the all-MA urn hypothesis against the half-MA urn hypothesis in the calibration experiment.

It is not clear how to apply this approach to the composite hypotheses $\theta \geq .791$ and $\theta < .791$, the composite hypotheses of interest in *Castaneda*.⁶⁵ Indeed, it is the case that the simple hypothesis $\theta = .791$ is supported over simple hypotheses about

64. This corresponds to a likelihood ratio of (approximately) 8.0×10^{142} .

65. Note that such a choice of hypotheses makes no real allowance for non-urn models as alternate hypotheses.

proportions very close to zero.⁶⁶ On the other hand, it has already been noted that the hypothesis $\theta = 339/870$ is supported over the hypothesis $\theta = .791$. If there are no grounds for *a priori* distinctions between different possible values of the proportion, then no general statement can be made about the hypotheses $\theta \geq .791$ and $\theta < .791$. One can point out, however, that the observation strongly supports proportions near $339/870$ over proportions near $.791$.⁶⁷ Thus, composite hypotheses restricting their attention to proportions near $339/870$ and $.791$ can be discussed using this approach.

4. *The Transition From Inference to Decision Making: Fisher's Significance Test*

At times, Fisher advocated something other than merely assessing the strength of evidence relative to a single hypothesis. According to Fisher, "it should be noted that the null hypothesis is never proved or established, but is possibly disproved in the course of experimentation. Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis."⁶⁸ Once again, we will approach composite hypotheses through simple hypotheses. As in the p-value approach, there is only a single hypothesis of interest. As will be seen below, however, this other Fisher approach can be said to have validation and even decision making aspects.

How does one perform one of Fisher's *significance tests* on a simple null hypothesis of the type at issue in *Castaneda*?⁶⁹ Roughly speaking, the procedure is as follows: one identifies a set of possible observations, call it R, which has a small probability if the null hypothesis is true. Thus, there is a high probability, assuming the null hypothesis is true, that an observation will not fall in R. One then carries out the experiment. If the observation falls in R, then the null hypothesis is rejected. For this reason, R is called the *rejection region*. If the observation does not fall in R, then the null hypothesis "survives." The "logic" of this procedure can be explained by the following. Recall from classical logic ("*modus*

66. The scale of Figure 1 should not deceive the reader. The likelihood curve does not "flatten." Rather, the likelihood starts at zero for $\theta = 0$, rises to its maximum for $\theta = 339/870$, and then descends to zero for $\theta = 1$. Even though the likelihood is vanishingly small for $\theta = .791$, the likelihoods are much smaller for θ 's close enough to zero.

67. For some general comments along these lines, see ROYALL, *supra* note 38, at 17-20.

68. R.A. Fisher, *Design of Experiments* 16 (8th ed. 1966).

69. There is a lack of uniformity on the terminology for the p-value and significance test approaches. See BARNETT, *supra* note 4, at 133.

tollens”) that if A implies E, then from not-E one can conclude not-A; on the other hand, if A implies E, then E does not allow one to conclude A. Fisher’s significance test procedure is the probabilistic analogue: if the null hypothesis implies WITH HIGH PROBABILITY that an observation will not fall in R, the fact that it does fall in R justifies rejecting the null hypothesis. On the other hand, the fact that an observation does not fall in R does not allow one to conclude that the null hypothesis is true.⁷⁰ Because there might be more than one set of observations R of a given probability, the force of this argument is stronger when the rejection region is chosen ahead of time.⁷¹

Significance tests have an evidential aspect. Their logic involves a certain type of evidential interpretation of the observations, and observations in rejection regions with smaller probabilities are said to represent stronger evidence against the hypothesis than those falling in rejection regions with larger probabilities.⁷² In this respect, the probability associated with R is called the *level of significance*. Moreover, because the rejection region is chosen ahead of time, the probability associated with R is the probability of mistakenly rejecting the null hypothesis when it is true.⁷³ In this respect, the probability associated with R is called *type I error*. By picking R, therefore, one can measure and control the probability of producing evidence leading to rejection of the null hypothesis when it is in fact true. Typically, one picks a “tail area” region that has a type I error close to, but not exceeding, one of the cutoff points used in the p-value qualitative yardstick.⁷⁴

Significance tests also have a validation function; one either rejects or does not reject the null hypothesis (i.e. the description). In this sense, one is making an assessment of the validity of the hypothesis itself. Note, however, that rejecting the null hypothesis means no more than that the hypothesis is rejected. One may have no specific alternative in mind to adopt. Moreover, not rejecting the null hypothesis does not mean that the null hypothesis has been accepted. Indeed, there may be other evaluations of the hypothesis yet to perform.

70. See ROYALL, *supra* note 38, at 72-73.

71. Cf. EDWARDS, *supra* note 60, at 177.

72. The reader is not alone if she wonders about an evidential approach that uses not only the observation that did occur, but also some that didn’t occur. See, e.g., ROYALL, *supra* note 38, at 69.

73. See, e.g., ROYALL, *supra* note 38, at 119-20.

74. See *supra* note 58 and accompanying text.

Viewed as a rejection procedure, a significance test has a prescriptive aspect. That is, certain actions usually are taken if the null hypothesis is rejected. In this sense, a significance test may be termed "decision making," although of a limited sort.

In the *Castaneda* context, if a typical social scientist were to perform a significance test for the hypothesis that the grand jury selection process is correctly modeled by an urn containing a proportion .791 of MA balls by looking at a sample of 870 selections, she would consider rejection regions with a type I error close to, but not exceeding, .05. For example, she might choose a rejection region consisting of an observed number of grand jurors less than 668, with a type I error of (approximately) .044.⁷⁵ After choosing the region, she would make the observation.⁷⁶

What about the composite null hypothesis $\theta \geq .791$? As with the p-value approach, one may deal with a composite null hypothesis in terms of the simple hypotheses contained therein. The rejection region above has probability less than .044 for any fixed proportion greater than .791. Because at most one proportion could be the correct proportion, significance test advocates would say that this region is an appropriate level .044 rejection region for a significance test of the composite null hypothesis $\theta \geq .791$.⁷⁷

5. *The Transition from Inference to Decision Making: The Neyman-Pearson Hypothesis Test*

The Neyman-Pearson approach is a multiple hypothesis approach that can be viewed as a complete shift from inference to decision making. In its traditional form, an experiment is performed to "choose" between competing hypotheses. Once again, simple hypotheses are the stepping-stones to composite hypotheses.

How does one perform a Neyman-Pearson *hypothesis test* on two simple hypotheses of the type at issue in *Castaneda*? Roughly speaking, the Neyman-Pearson procedure is as follows. In contrast to the law of likelihood approach, a null hypothesis is selected. As in the p-value and significance test approaches, the word "null" has

75. An R consisting of an observed number less than 669 has a probability of (approximately) .052.

76. Of course, this raises the question of what one does when the observations are made before the testing methodology is fully specified. Such a question is beyond the scope of this Article.

77. See R.A. FISHER, *STATISTICAL METHODS AND SCIENTIFIC INFERENCE* 49-50 (3d ed. 1972). Although *Castaneda* is not completely clear on the point, the Court seems to be using the p-value and significance tests approaches. 430 U.S. at 496 n.17.

a negative connotation here.⁷⁸ One then identifies a collection of possible observations, call it C . An experiment is performed. The null hypothesis is rejected in favor of the alternate hypothesis if the observation falls in C , and the null hypothesis is accepted if the observation does not fall in C . For this reason, C is called the *critical region*.

But how is C to be chosen? The basic idea is to choose C so that the resulting *decision procedure* is "good." In the traditional Neyman-Pearson framework, the notion of "goodness" is measured in terms of the two types of errors that one might make with respect to the null hypothesis — the *type I error*, which is the probability of mistakenly rejecting the null hypothesis if the null hypothesis is true; and the *type II error*, which is the probability of mistakenly accepting the null hypothesis if the null hypothesis is false, and the alternate simple hypothesis is true. But how is the choice of C to be made with respect to these two errors? Intuitively, there is some sort of tradeoff between the two error rates. For example, taking C to be the set of every possible observation yields a type I error of 1 and a type II error of 0, while taking C to be the empty set yields a type I error of 0 and a type II error of 1.

Suppose, for example, that one is going to draw five balls from an urn with an unknown proportion of MA balls. Suppose a decision is to be made between the null hypothesis that the proportion in the urn is .5 and the alternate hypothesis that the proportion is .75. Typically, one selects an upper bound for the type I error first and chooses, from among the regions yielding at most this type I error, a C that minimizes the type II error.⁷⁹ Suppose, for example, that one wants a type I error of at most $6/32$. Among the C 's with at most this probability under the null hypothesis, the one minimizing the type II error consists of 4 or 5 MAs. For this C , the type I error is $6/32$ and the type II error is $376/1024$.

What about the two composite hypotheses of interest in *Castaneda*; $\theta \geq .791$ and $\theta < .791$?⁸⁰ Again, composite hypotheses are handled by considering the simple hypotheses contained therein. Consider testing the simple hypothesis $\theta = .791$ against $\theta = \theta'$, where θ' is a fixed proportion less than .791. Suppose one wants

78. See, e.g., STEPHEN F. ARNOLD, *MATHEMATICAL STATISTICS* 288-89 (1990); BARNETT, *supra* note 4, at 170.

79. See, e.g., M.H. DEGROOT, *PROBABILITY AND STATISTICS* 449 (2d ed. 1986).

80. Note that the use of such an alternate makes no real allowance for non-urn models as alternate hypotheses.

a type I error close to, but not exceeding, .05. With the typical test procedure described above, standard calculations give a critical region C consisting of an observed number of Mexican-American grand jurors less than 668, with a type I error of (approximately) .044.⁸¹ Of course, the type II error depends on θ' .

What can we say about this C ? It is the case that C has a certain type I error property, no proportion greater than or equal to .791 results in a type I error greater than .044. It also can be shown that for any proportion less than .791, C yields the smallest type II error for any region satisfying the type I error property.⁸² As a result, Neyman-Pearson advocates would say that C is an appropriate .044 critical region for deciding between the null hypothesis $\theta \geq .791$ and the alternate hypothesis $\theta < .791$.⁸³

What exactly does "choosing" between the competing hypotheses mean? Many people say that hypothesis testing is validation. As one standard text puts it, "[t]he goal of a hypothesis test is to decide . . . which of two . . . hypotheses is true."⁸⁴ Jerzy Neyman, however, argued against this description of hypothesis testing.

The terms "accepting" and "rejecting" a . . . hypothesis are very convenient and are well established. It is important, however, to keep their exact meaning in mind and to discard various additional implications which may be suggested by intuition. Thus, to accept a hypothesis H means only to decide to take action A rather than action B . This does not mean that we necessarily believe that the hypothesis H is true. Also, if the application . . . "rejects" H , this means only that the rule prescribes action B and does not imply that we believe that H is false.⁸⁵

81. For a discussion of the mathematics involved, see, e.g., DEGROOT, *supra* note 79, at 442-49.

82. See, e.g., PAUL G. HOEL, SIDNEY C. PORT & CHARLES J. STONE, INTRODUCTION TO STATISTICAL THEORY 67-72 (1971).

83. See, e.g., *Id.*

84. LEE J. BAIN & MAX ENGELHARDT, INTRODUCTION TO PROBABILITY AND MATHEMATICAL STATISTICS 389 (2d ed. 1992) (emphasis omitted).

85. J. Neyman, First Course in Probability and Statistics 259 (1950). Indeed, an earlier example can be extended to show how a strict application of the Neyman-Pearson ideology makes even an evidential interpretation problematic. Suppose again that one is going to draw five balls from an urn with an unknown proportion of MA balls. Suppose a decision is to be made between the null hypothesis $\theta = .5$ and the alternate hypothesis $\theta = .75$. Consider the following two procedures. In the first procedure, one chooses five balls with a critical region of 2 MAs. With this critical region, the type I error is 10/32 and the type II error is 934/1024. The second procedure is a *randomized procedure*. One chooses five balls. If 5 MAs are

6. *Decision Making: Wald's Decision Theory*

For some, "statistics is concerned with decision making in the face of uncertainty."⁸⁶ In this sense, the Neyman view of hypothesis testing is a step in the right direction. As Abraham Wald noted, however, hypothesis testing suffers from several drawbacks as a general decision-making theory.⁸⁷ One problem is that the notion of the optimal course of action is somewhat limited, being restricted to the consideration of error rates alone, without taking into account the costs of the specific errors involved vis-à-vis the actions taken.

How might a frequentist describe the Neyman-Pearson procedure given in the preceding subsection for deciding between the null hypothesis $\theta \geq .791$ and the alternate hypothesis $\theta < .791$ using more general decision-theoretic terms?⁸⁸ Roughly speaking, the idea is as follows. The starting point is to characterize the *actions* available. In this case, there are two possible *actions*: (1) what is done if one is to accept the hypothesis $\theta \geq .791$; and (2) what is done if one is to accept the hypothesis $\theta < .791$.⁸⁹ Next, one defines a *loss function* that quantifies the loss associated with taking the available actions. In this case, for example, consider the

observed, the null hypothesis is rejected in favor of the alternate hypothesis. If 3 or fewer MAs are observed, the null hypothesis is accepted. If 4 MAs are observed, one generates a random number between 0 and 1. That is, assume a uniform distribution of the type pictured in Figure 4 *infra*. See, e.g., DEGROOT, *supra* note 79, at 103-05. One rejects the null hypothesis if the random number is less than or equal to .12. That is, one rejects 100% of the time on 5 MAs and 12% of the time on four MAs. Comparing the two procedures requires reconceptualizing the nonrandomized procedure so that it is a special case of the collection of randomized procedures; these details are beyond the scope of this Article. For a discussion, see ARNOLD, *supra* note 78, at 306-08. With the suitable theoretical framework, it can be shown that the randomized procedure has a type I error of .05 and a type II error of 732.4/1024. If one is interested only in error rates, there is no obvious reason not to prefer the randomized procedure. Yet there is an obvious evidential concern, how can one interpret the random number in terms of (the strength of) evidence for or against either of these hypotheses? This example casts doubt on inferential interpretations of hypothesis testing in general. For a discussion of some other "paradoxes" involving a strict application of Neyman-Pearson techniques, see Michael D. Perlman & Lang Wu, *The Emperor's New Tests*, 14 STAT. SCI. 355 (1999).

86. Herman Chernoff & Lincoln E. Moses, *Elementary Decision Theory* 1 (Dover ed. 1986).

87. See ABRAHAM WALD, *STATISTICAL DECISION FUNCTIONS V* (1950).

88. Some decision theorists do not like to use the word "hypothesis" because it incorrectly implies that the goal is to determine truth or falsity rather than to make decisions. See, e.g., JOHN W. PRATT, HOWARD RAIFFA & ROBERT SCHLAIFER, *INTRODUCTION TO STATISTICAL DECISION THEORY* 529 (1995).

89. Again, there is no real allowance here for non-urn procedures as alternate hypotheses.

loss function specifying a loss of 1 if the wrong hypothesis is accepted, and a loss of 0 otherwise.⁹⁰ Then, one defines a set of possible observations. Here, it is the number of Mexican-American grand jurors selected out of 870. One then defines a *decision procedure* that specifies the action to be taken upon any observation. In this example, the decision procedure is to accept $\theta \geq .791$ if there are at least 668 Mexican-American grand jurors and accept the hypothesis $\theta < .791$ otherwise.

The decision-theoretic approach begins by considering the *risk function* associated with a given decision procedure. In the *Castaneda* context, the risk function can be thought of as measuring the risk under the assumption that the selection process is correctly modeled by selecting balls from an urn with a given proportion of MA balls. More specifically, the *risk* of the procedure for a particular assumed-to-be-true proportion is obtained by "aggregating" the losses over all possible observations, where the aggregation is done by weighing the loss by the probability under the assumed proportion. With the Neyman-Pearson procedure described above, the following can denote the risk function:

$$R_{NP}(\theta') = \begin{cases} 1 \times P_{\theta = \theta'} \text{ (less than 668 in 870) if } \theta' \geq .791 \\ 1 \times P_{\theta = \theta'} \text{ (at least 668 in 870) if } \theta' < .791 \end{cases}$$

A frequentist sees this risk function as measuring the performance of the Neyman-Pearson procedure in the sense of giving the average loss over repeated applications of the procedure that occurs if the selection process is correctly modeled by selecting balls from an urn with proportion θ' .

It can be shown that the Neyman-Pearson decision procedure described above is *acceptable* or *admissible* in the sense that for any other decision procedure d with risk function R_d , if there is a θ' with $R_d(\theta') < R_{NP}(\theta')$, then there is a θ'' with $R_d(\theta'') > R_{NP}(\theta'')$.⁹¹ In most situations, nonadmissible decision procedures have little to suggest themselves in terms of the risk-evaluating decision-theoretic perspective.⁹²

There are, however, many admissible decision procedures here.⁹³

90. Gains would be indicated with negative numbers.

91. See, e.g., CHERNOFF & MOSES, *supra* note 86, at 338-39.

92. There are some situations in which they may be considered. See, e.g., GEORGE CASELLA & ROGER L. BERGER, *STATISTICAL INFERENCE* 480 (1990).

93. See, e.g., CHERNOFF & MOSES, *supra* note 86, at 338-39.

Is there anything else to recommend this one? Noting that the two probabilities appearing in the risk function represent type I and type II errors, one can describe an additional property of this risk function in terms of the error analysis of the Neyman-Pearson procedure discussed in the preceding subsection. Such a description, however, seems to have little to offer from a decision-theoretic perspective.⁹⁴

If there are no grounds for *a priori* distinctions between different possible values of the proportion, frequentists are largely left with intuitive criteria such as the *Minimax Property*.⁹⁵ Roughly speaking, a decision procedure satisfies the Minimax Property if it has the smallest possible *maximum risk*, where the maximum risk for a decision procedure is the "largest value" taken on by the risk function.⁹⁶ That is, one seeks to minimize the worst possible long run average loss.

In this example, there is a decision procedure that is admissible and minimax. It is a *randomized procedure*.⁹⁷ If the number of Mexican-American grand jurors selected is greater than 688, accept the hypothesis $\theta \geq .791$. If the number is less than 688, accept the hypothesis $\theta < .791$. If the number is 688, one generates a random number between 0 and 1, accepting the hypothesis $\theta < .791$ if the random number is less than or equal to (approximately) .767.⁹⁸ In this context with the 0-1 loss structure, the maximum risk for this procedure is $\frac{1}{2}$. Noting that the randomized procedure is roughly stated "accept $\theta \geq .791$ if the observed proportion of grand jurors selected is greater than or equal to .791 and accept $\theta < .791$ otherwise," the reader should not be surprised to discover that the procedure has some reasonable properties. The maximum risk for the Neyman-Pearson procedure is (approximately) .956.⁹⁹

C. *The Bayesians*

"[T]he Bayesian approach continues to gain adherents, but . . . there is still some way to go before we live in a fully

94. See, e.g., BARNETT, *supra* note 4, at 284-85.

95. *Id.* at 267.

96. Actually, the least upper bound of the values.

97. For a discussion of randomized procedures, see ARNOLD, *supra* note 78, at 306-08.

98. That is, assume a uniform distribution of the type pictured in Figure 4. See, e.g., DEGROOT, *supra* note 79, at 103-05.

99. This number is in fact the least upper bound of the risks and is obtained as θ approaches .791 from below.

Bayesian world."¹⁰⁰

1. Background

Some people assert that applying probability and statistics is about "individual degrees of belief," not some objective notion of the long run. According to Bruno de Finetti:

The abandonment of superstitious beliefs about . . . Fairies and Witches was an essential step along the road to scientific thinking. Probability, too, if regarded as something endowed with some kind of objective existence, is no less a misleading misconception, an illusory attempt to exteriorize our true probabilistic beliefs.

In investigating the reasonableness of our own modes of thought and behavior under uncertainty, all we require, and all that we are reasonably entitled to, is consistency among these beliefs, and their reasonable relation to any kind of relevant objective data ("relevant" in as much as subjectively deemed to be so). This is Probability Theory.¹⁰¹

The coin-flipping example again illustrates the parameters of such a position. Basically, a Bayesian talks about the probability of heads in terms of a particular, not a typical, observation. That is, the word "probability" refers to a specified situation or individual occurrence. Indeed, contrary to the frequentist view, it is not nonsense to say, "The probability is .5 that heads appeared the first time this coin was flipped." The sentence describes the speaker's degree of belief in this isolated historical event. According to this view of probability, individuals may disagree on the strengths of their beliefs, but they will agree on the mathematical rules obeyed, the most important of which is known as *Bayes' Theorem* (hence the term "Bayesian").¹⁰²

Bayesians assert that the mathematical rules allow us to update our beliefs in the light of new evidence. According to one Bayesian: "In broad outline, we take prior beliefs about various possible

100. Peter M. Lee, *Bayesian Statistics: An Introduction* viii (2d ed. 1997).

101. Bruno de Finetti, *THEORY OF PROBABILITY, A CRITICAL INTRODUCTORY TREATMENT*, VOL. 1, x (1970). For general discussions of such a position, see GILLIES, *supra* note 43, at 50-87; WEATHERFORD, *supra* note 46, at 219-42.

102. Although there is some debate about whether an approach emphasizing Bayes' Theorem requires a degree-of-belief notion of probability, this Article takes the increasingly prevalent position that it does. For a general discussion of the issue, see BARNETT, *supra* note 4, at 207, 242-49.

hypotheses and then modify these prior beliefs in the light of relevant data which we have collected in order to arrive at posterior beliefs."¹⁰³

This point of view is termed *subjective* in the sense that under this view the concept of "probability" involves consideration of a personal degree of belief. For this reason, frequentists and Bayesians have coexisted in a somewhat uneasy tension within the statistics community.¹⁰⁴

In the disciplinary terms of this Article, the debate between frequentists and Bayesians is less a debate about the scientific component of probability and statistics than it is a debate about the technological component. That is, there is less debate about the probability calculus *per se* than there is about the proper application.¹⁰⁵

Consider how one can abstract the *Castaneda* (under) representation issue so as to make it amenable to Bayesian analysis. For the purposes of this Article, the interest once again is in assertions that the selection process is correctly modeled in terms of selecting balls from an urn containing some proportion θ of balls labeled MA for Mexican-American. Bayesians assign degrees of belief to various assertions (i.e. hypotheses) about the proportion θ . The task is to update these degrees of belief in light of the evidence that there were 339 Mexican-American grand jurors out of 870 chosen.

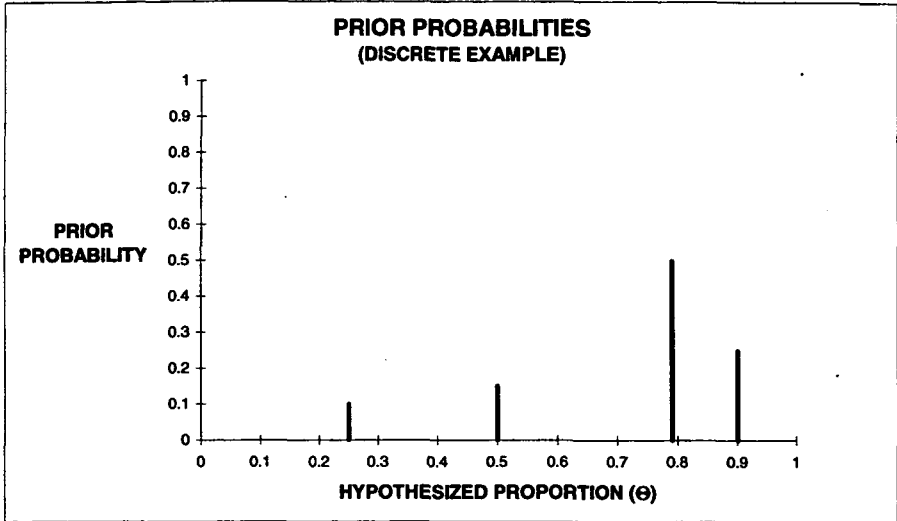
To illustrate the basic ideas, consider the universe of possible proportions to be finite. That is, consider what can be called a *discrete* approach. For example, one might assume that the possible proportions are .25, .5, .791, and .9. One assigns a *prior probability* or *prior degree of belief* to each of these proportions in such a way that the prior probabilities sum to 1. For example, one might assign the prior probability .1 to the assertion $\theta = .25$, the prior probability .15 to the assertion $\theta = .5$, the prior probability .5 to the assertion $\theta = .791$, and the prior probability .25 to the assertion $\theta = .9$. With reference to Figure 2, these prior probabilities can be pictured by a graph wherein the height of the line above the horizontal axis is the prior probability.

103. Lee, *supra* note 100, at 33.

104. See, e.g., ARNOLD, *supra* note 78, at 568-70.

105. See, e.g., BARNETT, *supra* note 4, at 72-73.

Figure 2



Using Bayes' Theorem, one updates these prior probabilities to produce *posterior probabilities* or *posterior degrees of belief* for these four proportions. In this context, Bayes' Theorem says that the posterior probability of a fixed proportion θ' given the evidence E is proportional to the product of (1) the likelihood ($P(E|\theta = \theta')$), and (2) the prior probability of the proportion ($P(\theta = \theta')$). Symbolically,

$$P(\theta = \theta'|E) = c \times P(E|\theta = \theta') \times P(\theta = \theta')$$

where "|" is read "given". The notation indicates that, as opposed to the frequentist perspective, the hypothesis itself has a probability attached to it. Thus, the likelihood is symbolized by $P(E|\theta = \theta')$, rather than $P\theta = \theta'(E)$. The constant of proportionality c is such that the posterior probabilities sum to 1. In this example,

$$P(\theta = .25|339 \text{ of } 870) = c \times P(339 \text{ of } 870|\theta = .25) \times P(\theta = .25),$$

$$P(\theta = .5|339 \text{ of } 870) = c \times P(339 \text{ of } 870|\theta = .5) \times P(\theta = .5),$$

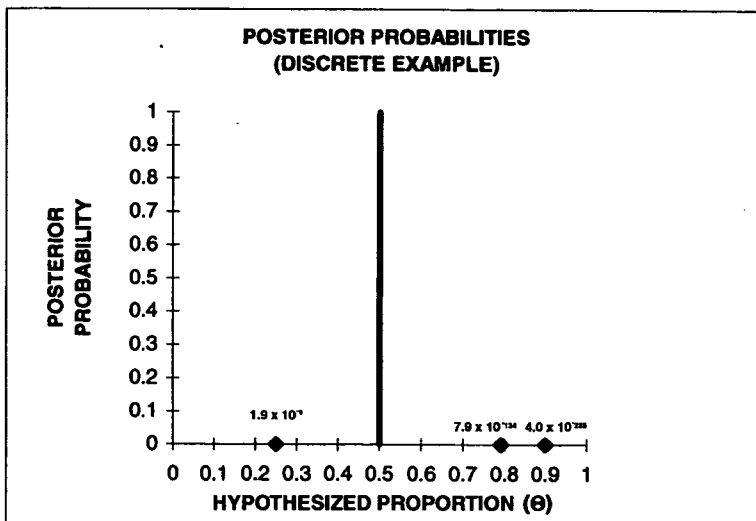
$$P(\theta = .791|339 \text{ of } 870) = c \times P(339 \text{ of } 870|\theta = .791) \times P(\theta = .791), \text{ and}$$

$$P(\theta = .9|339 \text{ of } 870) = c \times P(339 \text{ of } 870|\theta = .9) \times P(\theta = .9).$$

Here, $P(\theta = .25) = .1$, $P(\theta = .5) = .15$, $P(\theta = .791) = .5$, and $P(\theta = .9) = .25$. In this example, the extreme nature of the observation yields a posterior probability of very nearly 1 for the proportion .5, which is the proportion closest to the observed proportion 339/870. The posterior probabilities for the remaining proportions are vanishingly small.¹⁰⁶ One might consider it more reasonable to take the universe of possible proportions to consist of all proportions between 0 and 1. This can be called a *continuous* approach. Prior degrees of belief can be depicted with a graph. In such graphs, the height of the curve above a given proportion on the horizontal axis is not the prior probability for that proportion, rather the area under the curve above an interval is the prior probability that the proportion is in that interval. We call this curve the "density." In such graphs, the height of the curve above each proportion does not sum to 1, rather the area under the curve is 1. The following graph, Figure 4, represents a *uniform prior*.¹⁰⁷

106. Standard calculations yield: $P(\theta = .25 | 339 \text{ of } 870) \approx 1.9 \times 10^{-9}$, $P(\theta = .791 | 339 \text{ of } 870) \approx 7.9 \times 10^{-134}$, and $P(\theta = .9 | 339 \text{ of } 870) \approx 4.0 \times 10^{-286}$. Finally, $P(\theta = .5 | 339 \text{ of } 870)$ is 1 minus the sum of the other posterior probabilities, a number very close to 1! That is, the extreme nature of the evidence has put a "spike" at .5, the proportion closest to the observed proportion 339/870. These posterior probabilities may be pictured as follows in Figure 3 below:

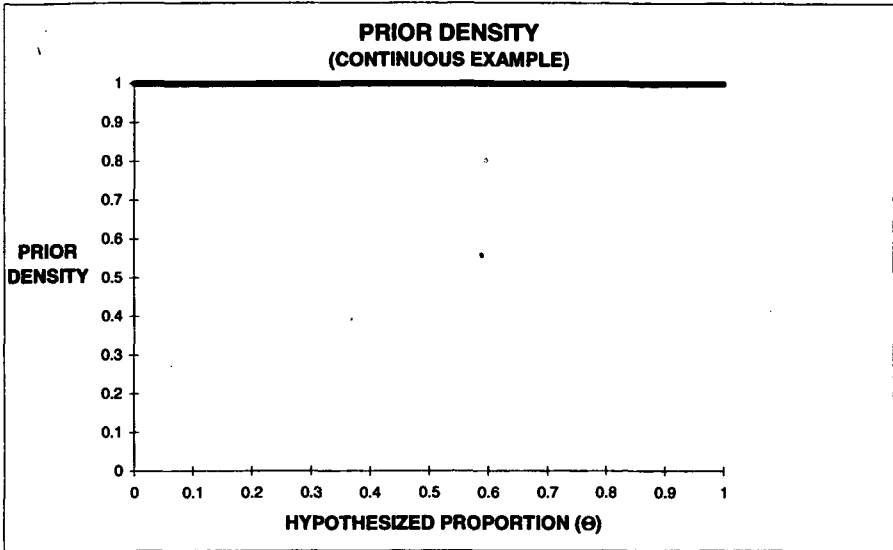
Figure 3



107. There is a voluminous debate about whether such a prior is "noninformative" in the sense that it represents some sort of "prior ignorance" that allows the evidence to speak to us unencumbered. This debate is not important for the purposes of this Article.

For general discussions, see JOSÉ M. BERNARDO & ADRIAN F.M. SMITH, *BAYESIAN THEORY*

Figure 4



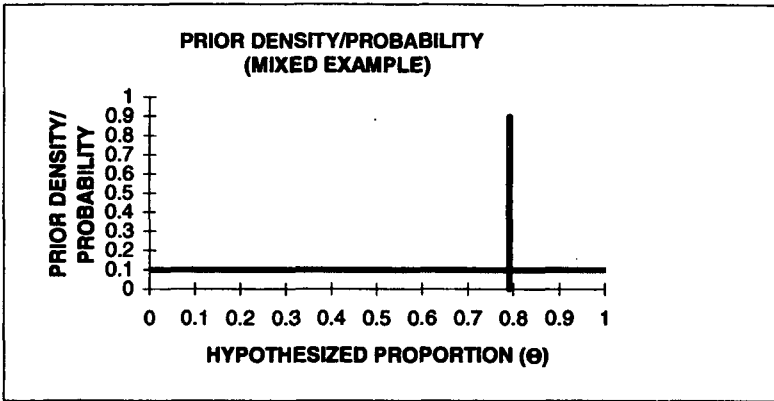
Now, the task is to update the graph using Bayes' Theorem. Standard calculations produce the following *posterior* graph (Figure 5) for the continuous prior beliefs given above in Figure 4.¹⁰⁸

357-67 (1994); LEE, *supra* note 100, at 83-86.

108. Note that the posterior is concentrated around the observed proportion 339/870. One might combine the two approaches.

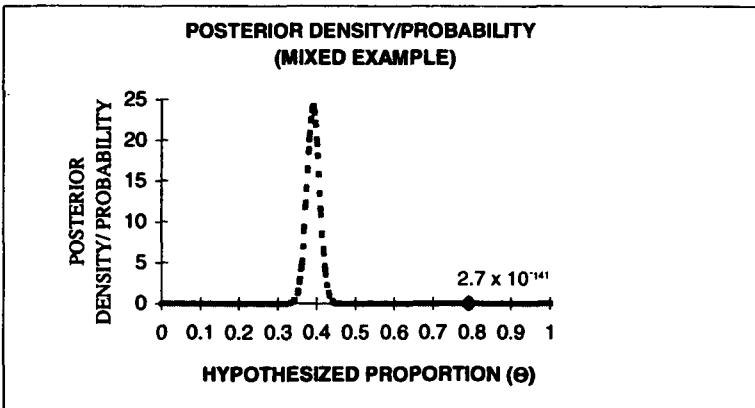
For example, one might assign a prior probability of .9 discretely to the single proportion .791, and assign the remaining probability to remaining proportions using the area approach, say in a uniform manner. One might depict the prior as follows:

Figure 6



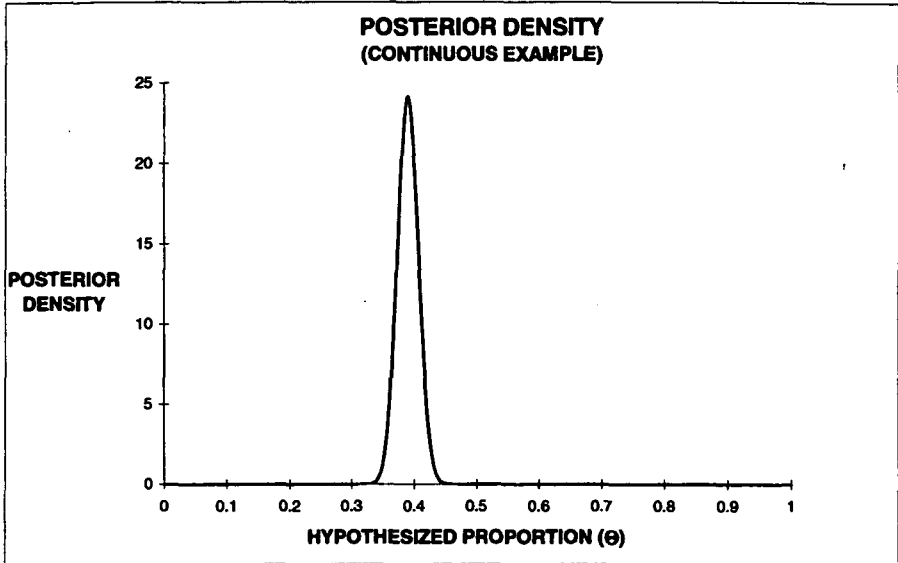
The posterior can be depicted as follow:

Figure 7



For a suggestion about how to use such a "spike and slab" approach in the jury discrimination context, see Stephen E. Fienberg & Joseph B. Kadane, *The Presentation of Bayesian Statistical Evidence in Legal Proceedings*, 32 THE STATISTICIAN 88, 94 (1983). In any case, none of the priors suggested here makes any real allowance for non-urn models as alternate hypotheses.

Figure 5



Bayesians do not agree on how far they should go to find analogues of frequentist approaches, especially to the extent that frequentist approaches stem from the need for indirection, and particularly with respect to hypothesis testing.¹⁰⁹ The Bayesian, unlike the frequentist, can make probability statements about individual hypotheses. The prior probabilities are degrees of belief about the hypotheses before evidence is introduced. Bayes' Theorem yields posterior degrees of belief about the hypotheses updated according to the evidence.¹¹⁰

The next subsections provide a brief introduction to some of the suggested Bayesian analogues for dealing with the composite hypotheses $\theta \geq .791$ and $\theta < .791$.

2. *Assessing the Strength of the Evidence with Respect to a Single Hypothesis: Bayesian P-values*

Some Bayesians suggest that the Bayesian measure of the strength of the evidence with respect to a single hypothesis is given by the posterior probability of that hypothesis.¹¹¹ In the *Castaneda*

109. For an introduction to the issue, see LAWRENCE D. PHILLIPS, *BAYESIAN STATISTICS FOR SOCIAL SCIENTISTS* 333-34 (1973).

110. Note that simple hypotheses cannot be analyzed under the continuous approach because any specific proportion has probability zero. The spike and slab approach will have the same problem to the extent it is continuous.

111. See, e.g., CASELLA & BERGER, *supra* note 92, at 401. Some feel this assertion makes

context, the obvious null hypothesis of interest is $\theta \geq .791$. In the discrete example above, the posterior probability for this hypothesis is the sum of the posterior probabilities for the proportions .791 and .9. As indicated above, this sum is a vanishingly small number.¹¹²

In the continuous example, it is the area under the posterior curve (see Figure 5) to the right of the proportion .791, again a vanishingly small number.¹¹³

3. *Assessing the Strength of the Evidence with Respect to Two Hypotheses: Bayes Factors*

Some Bayesians talk about assessing the strength of evidence with respect to two hypotheses. Recall that for simple hypotheses A and B and observation E, the frequentist takes the likelihood ratio, $P_A(E)/P_B(E)$, to measure the strength of the evidence with respect to the two hypotheses. As seen, however, with no grounds for *a priori* distinctions between different possible values of the proportion, it is difficult for the frequentist to generalize the law of likelihood approach to composite hypotheses of the type at issue in *Castaneda*. In the Bayesian context, however, we may compute the ratio: $P(339 \text{ in } 870 | \theta < .791) / P(339 \text{ in } 870 | \theta \geq .791)$.

It turns out that this ratio is equal to the ratio of the posterior odds ratio to the prior odds ratio. That is, it is equal to:

$$\frac{P(\theta < .791 | 339 \text{ in } 870) / P(\theta \geq .791 | 339 \text{ in } 870)}{P(\theta < .791) / P(\theta \geq .791)}$$

Either of these two equivalent ratios is called the Bayes Factor.¹¹⁴

Some Bayesians assert that the Bayes Factor assesses the strength of the evidence with respect to the two hypotheses because it provides a measure of the extent to which the observed data have increased or decreased the relative degrees of belief.¹¹⁵

more sense in the situation of a noninformative prior as then the posterior degree of belief is "solely dependent" on the evidence. See *supra* note 107. Indeed, in certain such situations the posterior probability will be numerically equal to the frequentist p-value. See PRATT, RAIFFA & SCHLAIFER, *supra* note 88, at 533-34; LEE, *supra* note 100, at 122.

112. This sum is (approximately) $7.9 \times 10^{-134} + 4.0 \times 10^{-285}$.

113. This area is less than 10^{-142} .

114. See, e.g., BARNETT, *supra* note 4, at 216-17. Some Bayesians refer to the logarithm of the Bayes Factor as the *weight of evidence*. *Id.*

115. See, e.g., BERNARDO AND SMITH, *supra* note 107, at 390. It turns out that the Bayes Factor for composite hypotheses involves weighted averages of the likelihoods of the proportions making up the two hypotheses, where the weights depend on the prior. Cf. LEE, *supra* note 100, at 119-20. Because of this dependence on the prior, some Bayesians have trouble using this as an assessment based solely on the strength of the evidence unless the

In both the discrete and continuous examples above, the Bayes Factor is a very large number, indicating that the evidence favors $\theta < .791$.¹¹⁶

4. Bayesian Significance Tests

Significance testing may be done in the Bayesian framework. In the *Castaneda* context, the obvious null hypothesis of interest is $\theta \geq .791$. A Bayesian significance test, analogous to the frequentist significance test, can be performed by rejecting this hypothesis if the posterior probability is small enough, according to some predetermined criterion.¹¹⁷ As discussed above, the posterior probability of the hypothesis $\theta \geq .791$ is vanishingly small in both the discrete and continuous examples.

5. Bayesian Hypothesis Tests

Hypothesis testing may also be done in the Bayesian framework. Suppose the null hypothesis $\theta \geq .791$ is to be tested against the alternate hypothesis $\theta < .791$. With the Bayesian methodology, the posterior probabilities of these hypotheses can be directly computed. Several strategies are possible.¹¹⁸ Treating the hypotheses symmetrically, one could decide to accept the hypothesis with the greater posterior probability, singling out the null hypothesis only for ties. More in line with the frequentist approach, one could single out the null hypothesis and accept the alternate hypothesis if the posterior probability of the null hypothesis is small according to some predetermined criterion. As discussed above, the posterior probability of the hypothesis $\theta \geq .791$ is vanishingly small in both the discrete and continuous examples, from which it follows that the posterior probability of

Bayes Factor is "relatively little affected within reasonable limits" by changes in the weights. LEE, *supra* note 100, at 120. See also *supra* notes 107, 111. In the discrete example, the Bayes Factor for two simple hypotheses turns out to be equal to the likelihood ratio endorsed by the frequentists. See, e.g., LEE, *supra* note 100, at 119.

116. In the discrete example given above, the Bayes Factor for $\theta < .791$ versus $\theta \geq .791$ is (approximately) 3.8×10^{133} . In the continuous example, the Bayes Factor for $\theta < .791$ versus $\theta \geq .791$ is at least 10^{141} .

117. See BARNETT, *supra* note 4, at 214-15; BERNARDO & SMITH, *supra* note 107, at 413. Other approaches have been suggested. See *id.* at 475-76. In any case, significance testing raises the issue of the prior to the extent that such testing is seen as evidential. See *supra* notes 107, 111, and 115.

118. See CASELLA & BERGER, *supra* note 92, at 354-55. Although this is dependent on the prior, there is no evidential issue if inference has been eschewed altogether in favor of decisionmaking. See *supra* notes 107, 111, 115, and 117.

the alternate is very close to 1 in both examples.¹¹⁹

6. *Bayesian Decision Theory*

Consider again the two hypotheses $\theta \geq .791$ and $\theta < .791$. As noted above, frequentist decision theorists are led to things like the Minimax Property when they have no *a priori* grounds for distinguishing among various values of θ .¹²⁰ The Bayesian perspective offers other avenues for distinguishing among decision procedures. The Bayesian can, for example, use the prior degrees of belief as weights to find the average (i.e. expected) value of the risk over all possible values of the proportion. For example, the expected risk for the Neyman-Pearson procedure in the discrete example is:

$$R_{NP}(.25) \times P(\theta = .25) + R_{NP}(.5) \times P(\theta = .5) + R_{NP}(.791) \times P(\theta = .791) + R_{NP}(.9) \times P(\theta = .9), \text{ which is (approximately) } .022.$$

This suggests the *Bayes Decision Principle*: use a decision procedure that minimizes the expected risk.¹²¹ Assuming the 0-1 loss structure described in the frequentist analysis, the resulting decision procedure is to accept the hypothesis with the larger posterior probability.¹²² In the discrete example, the procedure is to accept $\theta < .791$ if the number of Mexican-American grand jurors is less than 570, and accept $\theta \geq .791$ otherwise. The expected risk for this procedure is vanishingly small.¹²³

D. *What Does This Tell us About the Concept of Law as a Discipline?*

Like perspectives, which rightly gazed upon Show nothing but confusion, eyed awry Distinguish form.¹²⁴

No doubt many mathematicians and legal scholars will complain about the discussion presented so far. At the descriptive level, mathematicians will point to a lack of breadth (e.g. parameter estimation) and depth (e.g. two-sided tests). Moreover, at the

119. In the discrete example, the posterior probability of the hypothesis $\theta \geq .791$ is (approximately) $7.9 \times 10^{-134} + 4.0 \times 10^{-285}$, from which it follows that the posterior probability of the hypothesis $\theta < .791$ is (approximately) $1 - (7.9 \times 10^{-134} + 4.0 \times 10^{-285})$. In the continuous example, the posterior probability of the hypothesis $\theta < .791$ is less than 10^{-142} , from which it follows that the posterior probability of the hypothesis $\theta < .791$ is at least $1 - 10^{-142}$.

120. See BARNETT, *supra* note 4, at 267.

121. *Id.* at 267-68. Again, some provision will have to be made for ties.

122. See CASELLA & BERGER, *supra* note 92, at 472-78.

123. It is less than 10^{-20} .

124. Richard II, Act II, Scene ii.

normative level, there is no discussion of the general frequentist and Bayesian schools or of attempts to reconcile them, nor is there any extensive evaluation of the various approaches contained within the schools.¹²⁵ Legal scholars might have similar complaints.

For the purposes of the type of crossdisciplinary education considered in this Article, however, the presentation is more than sufficient. Remember that the primary goal is for students to think about law as a discipline, not to become intelligent consumers of various mathematical techniques. It is difficult for students to step back and look at law as a discipline when there is no "back." The idea of the type of crossdisciplinary education described here is to provide students with another vantage point.

Considering mathematics as a discipline offers an opportunity for students to think about law as a discipline. Consider, then, mathematics as a science, an art, and a technology.¹²⁶ As a science, mathematics deals with objects whose essential properties involve number, shape, and function.¹²⁷ The heart of mathematical classification is the definition-theorem-proof method.¹²⁸ As an art, mathematics interprets using number, shape, or function in conjunction with a medium that consists of a highly refined symbolic language.¹²⁹ Finally, mathematics as a technology is used

125. The reader can start with BARNETT, *supra* note 4; GILLIES, *supra* note ; ROYALL, *supra* note 38; WEATHERFORD, *supra* note 46.

126. See MORRIS KLINE, *MATHEMATICS: A CULTURAL APPROACH* 1-10 (1962).

127. See NATIONAL RESEARCH COUNCIL BOARD ON MATHEMATICAL SCIENCES, *MATHEMATICAL SCIENCES: SOME RESEARCH TRENDS* 21 (1988).

128. See, e.g., Y. MANIN, *A COURSE IN MATHEMATICAL LOGIC* 48 (1977) (stating "the ideal for what constitutes a mathematical demonstration of a 'nonobvious truth' has remained unchanged since the time of Euclid; we must arrive at such a truth from 'obvious' hypotheses, or assertions which have already been proved, by means of a series of explicitly described, 'obviously valid' elementary deductions"). This is the "ideal," but the required rigor has changed from time to time. See Wilder, *Relativity of Standards of Mathematical Rigor*, 3 *DICTIONARY OF THE HISTORY OF IDEAS* 170 (1973); see also Grabiner, *Is Mathematical Truth Time-Dependent?*, 81 *AMERICAN MATHEMATICAL MONTHLY* 354 (1974).

129. See Henri Poincaré, *The Relations of Analysis and Mathematical Physics*, 4 *BULL. AM. MATHEMATICAL SOC'Y* 247, 248 (1898) (stating "[Mathematics has] an end esthetic [A]depts find in mathematics delights analogous to those that painting and music give. They admire the delicate harmony of number and of forms; they are amazed when a new discovery discloses for them an unlooked for perspective . . ."). For other discussions of mathematics as an art, see NATHAN A. COURT, *MATHEMATICS IN FUN AND EARNEST* 127-40 (1964); P.R. Halmos, *Mathematics as a Creative Art*, 56 *AM. SCIENTIST* 375 (1968); J.W.N. Sullivan, *Mathematics As an Art*, in *ASPECTS OF SCIENCE: SECOND SERIES* 80 (1926); Henri Poincaré, *Mathematical Creation*, *SCI. AM.*, Aug. 1948, at 54. See also SCOTT BUCHANNAN, *POETRY AND MATHEMATICS* (1929); JERRY P. KING, *THE ART OF MATHEMATICS* (1992). Many mathematicians exalt this dimension of mathematics above all others. See Lynn A. Steen, *Mathematics Today*, in *MATHEMATICS TODAY* 1, 10 (Lynn A. Steen ed., 1978) (stating "beauty and elegance have more to do with the value of a mathematical idea than does either strict truth or

to solve problems that can be modeled in terms of number, shape, and function.¹³⁰ Mathematics can be used as a kind of "mirror" to examine the scientific, artistic, and technological components of law. The goal should be less to develop a full picture of law as a discipline, than to impart to students an understanding of what such an enterprise entails.

For example, students can consider law as an art. One of the artistic aspects of mathematics involves the idea of the "elegant" or "ingenious" proof.¹³¹ One can begin a crossdisciplinary course of the type envisioned here with an axiomatic treatment of probability, and students can be asked to do proofs! Homework will provide examples where students can be asked to compare their solutions along these artistic lines. Students can then consider a similar situation in law, the "elegant" or "ingenious" opinion. An earlier article hints how this might be done using Cardozo's well-known *Allegheny College* opinion.¹³²

As an example of what might be done to get students to think about law as a science, students can be asked whether they find the standard casebook "note problems" any more or less difficult than the "word problems" in the quantitative methods course, and if so, why.

Instead of considering the components of law individually, students can think about the relation between them. The frequentist-Bayesian divide presents an excellent opportunity for students to think about the relation between law as a science and law as a technology. The scientific aspect of probability and statistics is encompassed by the axiomatic study of certain sets and functions. As mentioned above, the frequentist-Bayesian debate is less an argument about the axiomatics than it is a fundamental disagreement about the proper technological stance associated with such axiomatic systems. The mathematician A.N. Kolmogorov, who

possible utility."); see also G.H. HARDY, A MATHEMATICIAN'S APOLOGY (3d prtg. 1967).

130. For a general discussion with a number of interesting examples, see Felix E. Browder & Saunders MacLane, *The Relevance of Mathematics*, in MATHEMATICS TODAY 323 (Lyin A. Steen ed., 1978).

131. See generally, WILLIAM DUNHAM, JOURNEY THROUGH GENIUS: THE GREAT THEOREMS OF MATHEMATICS (1990).

132. See *supra* text accompanying note 27. Another aspect of mathematical elegance involves parsimony in the choice of axioms. See, e.g. RAYMOND L. WILDER, EVOLUTION OF MATHEMATICAL CONCEPTS 9 (1968). In fact, axiom systems for probability provide a good illustration, and the point can be emphasized in class as the presentation unfolds. Students can then be asked to think about similar situations in law. For example, students can be asked to think about the "mailbox rules" in the contract Restatements.

provided one of the earliest axiomatic treatments of probability, was well aware of the basic distinction between science and technology in this area:

The theory of probability, as a mathematical discipline, can and should be developed from axioms in exactly the same way as Geometry and Algebra. This means that after we have defined the elements to be studied and their basic relations, and have stated the axioms by which these relations are to be governed, all further exposition must be based exclusively on these axioms, independent of the usual concrete meaning of these elements and their relations.¹³³

Indeed, students can study the axiomatics independent of any discussion of technological interpretations. Students can be told that the usual sorts of problems involving dice and cards are stated so as to provide a shorthand for various complicated sets and functions. After students have worked through the basic framework, the frequentist-Bayesian divide can be presented. This provides an excellent example of the stark distinctions that can be drawn between the scientific and technological components of a discipline. On the other hand, the dice and card examples can be reintroduced in pointing out that the development of the field of probability is due largely to the investigation of a number of practical problems.¹³⁴ This reminds students that the scientific and technological components of a discipline often are symbiotic and cannot be separated too neatly. With this background, students can be asked about the relation between law as a science and law as a technology.

Students also can think about the current evolution of law as a discipline. Although opinion on Bayesian methods has swung back and forth over the past two centuries, the recent interest in Bayesian techniques is traceable in part to the general availability of modern computing power.¹³⁵ Assignments utilizing statistical software will make the impact of computing power clear. Students can be asked to consider the effect of legal databases on law as a discipline. For example, according to one legal scholar:

133. A.N. KOLMOGOROV, *FOUNDATIONS OF THE THEORY OF PROBABILITY* 1 (1956).

134. See IAN HACKING, *THE EMERGENCE OF PROBABILITY* 11-12 (1984).

135. See, e.g., GEORGE E.P. BOX & GEORGE C. TIAO, *BAYESIAN INFERENCE IN STATISTICAL ANALYSIS* 1 (1973); cf. ANDREW GELMAN, JOHN B. CARLIN, HAL S. STERN & DONALD B. RUBIN, *BAYESIAN DATA ANALYSIS* 3-4 (1995).

To use the computer, we have to think like a computer. The more we think in a highly simplified, individual-word fashion, the more effective our computer searches will be. But the more we think in computerese, the further we move from our traditional way of "thinking like a lawyer."¹³⁶

Students also can discuss specific legal doctrine. *Castaneda* suggests some interesting questions about the prima facie case in discrimination suits. In making out the prima facie case in *Castaneda*, plaintiff was not required to adjust the statistical analysis for statutorily prescribed qualifications,¹³⁷ nor was plaintiff required to present a separate analysis for other possible time frames.¹³⁸ In the terminology of this Article, the prima facie case has a scientific aspect as a type of classification device. A prima facie case results in a preliminary classification that will become permanent in the absence of a rebuttal. However, the prima facie case also has a technological aspect. For example, in terms of allocating relative burdens, the elements of a prima facie case might depend on issues such as which party is in a better position to maintain and bring forth certain types of data.¹³⁹ In this sense, there is an interesting contrast with general statistical practice. If plaintiff's statistical brief were to be turned into an analysis for a statistics journal, he would be required to consider the adjustments described above before making any conclusion. Indeed, the unavailability of relevant data for making the adjustments might be a fatal blow to publication. Given this contrast, students can be asked to think about the nature and use of the prima facie case in discrimination suits.¹⁴⁰

136. Barbara Bintliff, *From Creativity to Computerese: Thinking Like a Lawyer in the Computer Age*, 88 LAW LIB. J. 338, 346 (1996). See also Susan W. Brenner, PRECEDENT INFLATION (1992); Carol Bast & Ransford Pyle, *Legal Research in the Computer Age: A Paradigm Shift?*, 93 LAW LIB. J. 285 (2001); Robert C. Berring, *Collapse of the Structure of the Legal Research Universe: The Imperative of Digital Information*, 69 WASH. L. REV. 9 (1994); Robert C. Berring, *Legal Research and Legal Concepts: Where Form Molds Substance*, 75 CALIF. L. REV. 15 (1987).

137. These qualifications were that the grand juror be a citizen of Texas and the county, be of sound mind and good moral character, be literate, have no prior felony conviction, and be under no pending indictment or other legal accusation for theft or of any felony. 430 U.S. at 484.

138. Plaintiff considered an 11-year time frame. The State District Commissioner at the time of plaintiff's indictment had been in office only 2 ½ years. 430 U.S. at 496 n.16.

139. See Sara Sun Beale, *Integrating Statistical Evidence and Legal Theory to Challenge the Selection of Grand and Petit Jurors*, 46 LAW & CONTEMP. PROBS. 269, 272-73 (1983).

140. For example, students might start with something like Kingsley R. Browne,

The different mathematical methodologies described in Part III suggest several questions. Is the classification in Castaneda "objective" or "subjective"? Put perhaps a second way, is the classification involved in *Castaneda* about "facts," "beliefs about facts," or something else? Put perhaps a third way; is "proof," "persuasion," or something else the appropriate cognate as it relates to *Castaneda*?¹⁴¹ The frequentist-Bayesian divide illustrates the potential importance of these questions. Even assuming a cognate with mathematical significance, is the ultimate job inference or decisionmaking?¹⁴² If inference, is it about assessing (the strength of) evidence (with respect to hypotheses) or about assessing (the validity of) hypotheses? Is there only a single hypothesis of interest, or are two hypotheses to be considered? Do the answers depend on the prima facie context? The p-value, law of likelihood, significance test, hypothesis test, and general decision-theoretic approaches indicate the potential importance of these questions. Finally, the Bayesian approach suggests questions about the exact nature of the presumption of innocence.¹⁴³ This list of questions is meant to be illustrative, not exhaustive. Others will no doubt suggest themselves to the reader.

It is not suggested that all students take a quantitative methods course, but law schools should consider developing a range of "capstone" courses at the crossdisciplinary, interdisciplinary, and transdisciplinary levels. In any case, legal education should encourage students to think along such lines throughout their careers.

IV. CONCLUSION

Many jobs that now require a college degree do so only out of

Statistical Proof of Discrimination: Beyond Damned Lies, 68 WASH. L. REV. 477 (1993).

141. For example, students might start with something like Ronald J. Allen, *The Nature of Juridical Proof*, 13 CARDOZO L. REV. 373, 393-96 (1991); Stephen E. Fienberg & Mark J. Schervish, *The Relevance of Bayesian Inference for the Presentation of Statistical Evidence and for Legal Decisionmaking*, 66 B.U.L. REV. 771 (1986); Craig R. Allen, *Some Limits on the Use of Bayesian Theory in Evidence Law*, 57 IND. L.J. 1 (1982).

142. For example, students might start with something like John Kaplan, *Decision Theory and the Factfinding Process*, 68 STAN. L. REV. 1065 (1968).

143. For example, students might start with something like Lawrence Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1368-72 (1971); Lawrence Tribe, *A Further Critique of Mathematical Proof*, 84 HARV. L. REV. 1810, 1815-17 (1971); Michael O. Finkelstein & William B. Fairley, *The Continuing Debate over Mathematics in the Law of Evidence*, 84 HARV. L. REV. 1801, 1807-09 (1971); Fienberg & Kadane, *supra* note 108, at 92-93.

professional tradition or expectations, rather than an inherent need for four years or more of higher education. Lawyers are a good example. Most of a lawyer's skills can be obtained from a paralegal school and a good high school forensics class. This would be more cost-effective than seven years of higher education¹⁴⁴

This Article is part of a work in progress. The goal of this project is to illustrate how one might consider law as a discipline. There is no claim that the approach sketched in this sequence of articles is the only approach, but American law schools must do more in this regard. One reason is obvious. Such a consideration is the sign of a mature subject area. There is, however, another important reason. In today's world, traditional education providers (i.e. colleges and universities) are facing increasing competition from nontraditional rivals. Education providers who do not think carefully about the nature of their offering will not be able to compete with those who do. Simply put, serious thinking by the current professoriate about law as a discipline may be a matter of survival.

144. Andrew Peyton, *Wrench-Wrestling vs. Pencil-Pushing*, THE AMERICAN ENTERPRISE, 58 Sept./Oct. (1996).