

Systematic Content Analysis of Judicial Opinions

Mark A. Hall†
Ronald F. Wright††

*Professor Herman Oliphant, in his 1928 inaugural address as President of the American Association of Law Schools:*¹

Our case material is a gold mine for scientific work. It has not been scientifically exploited. . . . We should critically examine all the methods now used in any of the social sciences and having any useful degree of objectivity.

INTRODUCTION

THE RESTATEMENT (FIRST) OF CODING CASES

Legal scholars, the mockingbirds of the academy, are great borrowers of scholarly methods. We experiment with the tools of historians, economists, sociologists, literary theorists, moral philosophers, and others, often to great effect. Yet despite these innovative efforts to study legal doctrines and institutions through different lenses, legal scholars have yet to identify their own unique *empirical* methodology. Instead, empirical legal methods are often standard applications of basic social science methods to subjects of (sometimes trifling) legal interest. In doing this kind of work, law professors may ably step

Copyright © 2008 California Law Review, Inc. California Law Review, Inc. (CLR) is a California nonprofit corporation. CLR and the authors are solely responsible for the content of their publications.

† Fred D. & Elizabeth L. Turnage Professor of Law, Wake Forest University School of Law and School of Medicine. My work on this paper was supported by a grant from the Robert Wood Johnson Foundation through its Investigator Awards in Health Policy Research program. I am especially grateful to Sally Shumaker, who first encouraged me to use content analysis.

†† Associate Dean for Academic Affairs and Professor of Law, Wake Forest University. Many people gave us useful information and insights, but we are especially grateful to Peter Jacobson, Jason Czarnecki, Jim Brudney, and Polk Wagner, and for comments received at the first annual Conference on Empirical Legal Studies and at a Duke Law School faculty workshop. Daniel Finegan provided able research assistance, and Ellen Makaravage was instrumental in locating publications.

1. Herman Oliphant, *A Return to Stare Decisis*, 14 A.B.A. J. 71, 161 (1928).

into the shoes of social scientists,² but their methods are not uniquely or especially legal methods. They have only a weak claim to engaging in a distinct disciplinary approach. Social scientists trained in disciplines other than law can do what empirical legal scholars do equally well or better.

We propose in this Article that one standard social science technique—content analysis—could form the basis for a uniquely legal empirical methodology. On the surface, content analysis appears simple, even trivial, to some. Using this method, a scholar collects a set of documents, such as judicial opinions on a particular subject, and systematically reads them, recording consistent features of each and drawing inferences about their use and meaning.³ This method comes naturally to legal scholars because it resembles the classic scholarly exercise of reading a collection of cases, finding common threads that link the opinions, and commenting on their significance.⁴ But content analysis is more than a better way to read cases. It brings the rigor of social science to our understanding of case law, creating a distinctively legal form of empiricism.

We are not proselytizers of this method. It has certain advantages, along with substantial limitations, compared to conventional legal analysis. What we claim is that, when one reads cases this way, one engages in a uniquely legal empirical method—a way of generating objective, falsifiable, and reproducible knowledge about what courts do and how and why they do it. Econometricians, political scientists, or sociologists can use these or other methods to study questions that *relate to* the law, such as external *causes and effects* of the law and legal institutions. But content analysis aims for a scientific understanding of the *law itself* as found in judicial opinions and other legal texts, a subject matter that plays to the strengths of legal scholars. While one early advocate of this method claims that it points us directly toward the Holy Grail of a century-long quest for a true legal science, our view is not this grandiose.⁵ Instead, we

2. For a debate about how well legal scholars have applied standard social science methods, see, for example, Frank Cross et al, *Above the Rules: A Response to Epstein and King*, 69 U. CHI. L. REV. 135 (2002); Lee Epstein & Gary King, *The Rules of Inference*, 69 U. CHI. L. REV. 1 (2002); Michael Heise, *The Past, Present, and Future of Empirical Legal Scholarship: Judicial Decision Making and the New Empiricism*, 2002 U. ILL. L. REV. 819; Richard L. Revesz, *A Defense of Empirical Legal Scholarship*, 69 U. CHI. L. REV. 169 (2002); Gregory C. Sisk & Michael Heise, *Judges and Ideology: Public and Academic Debates About Statistical Measures*, 99 NW. U. L. REV. 743, 791-793 (2005).

3. See KLAUS KRIPPENDORFF, *CONTENT ANALYSIS: AN INTRODUCTION TO ITS METHODOLOGY* 18 (2d ed., 2004) (defining content analysis as “a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use”).

4. Cf. David Zaring, *The Use of Foreign Decisions by Federal Courts: An Empirical Analysis*, 3 J. EMPIRICAL LEGAL STUD. 297, 303 (2006) (suggesting that citation analysis “depends on the sort of doctrinal analysis that legal scholars are particularly suited to do”).

5. Reed C. Lawlor, *Personal Stare Decisis*, 41 S. CAL. L. REV. 73 (1967). For the beginnings of this quest, see, for example, Roscoe Pound, *Mechanical Jurisprudence*, 8 COLUM. L. REV. 605 (1908) (calling for a more “exact justice, that is for a justice whose operations within

maintain that content analysis makes legal scholarship more consistent with the basic epistemological underpinnings of other social science research. The method combines a disciplined focus on legal subject matter with an assumption that other investigators should be able to replicate the research results. Put another way, the research results matter more than the researcher's authority.

We began this enterprise vaguely aware that several legal scholars in recent years had used content analysis. Once we collected and read these studies, however, a rich history came into focus. As we discuss in Part I, content analysis first appeared decades ago, and the method grew in fits and starts. Legal scholars developed their uses of content analysis organically, similar to the way that judges develop the common law. Many legal content analysts designed their studies without referring to other examples of this method, and only in retrospect did a set of methodological principles start to emerge. This type of "common law" development created an adaptive form of content analysis, one that is partially based on techniques prescribed in other social sciences, but which differs from the classic approach in some respects. The distinctive concerns of lawyers and legal academics have determined which standard techniques are used and which are neglected. The sporadic ad hoc growth of this methodology calls out for a systematic analysis of its uses and tenets.

Like common law judges, academic content analysts have been slow to generalize about their choices. Users note that the technique adds something to the traditional interpretive enterprise of reading a few appellate opinions and commenting on their themes and likely social effects. For one thing, they say, content analysis allows the researcher to deal with larger numbers of cases, which provides a truer measure of broad patterns in the case law. The method also helps a researcher to sort out the interaction of multiple factors that bear on an outcome in the legal system. Beyond such passing observations, however, academics have not thought systematically about what content analysis adds to traditional legal analysis.

In Part II, we survey the questions that scholars have tried to answer through content analysis, and, using that experience, we generalize about the technique's strengths and weaknesses. Which questions about the law and legal institutions are well suited to this method and which are not? What does a scholar gain or lose by employing this method rather than traditional interpretive legal scholarship? We argue that the uses of content analysis that best combine the strengths of the legal scholar and the social scientist ask about the internal interaction of facts and arguments in an opinion. Content analysis

reasonable limits, may be predicted in advance of action"). See generally BENJAMIN N. CARDOZO, *THE PARADOXES OF LEGAL SCIENCE* (Greenwood Press 1970) (1928), reprinted in *SELECTED WRITINGS OF BENJAMIN NATHAN CARDOZO* 253 (M. Hall ed., 1947); Oliver Wendell Holmes, *Law in Science and Science in Law*, in *COLLECTED LEGAL PAPERS* 210, 242 (1921).

also works best when the judicial opinions in a collection hold essentially equal value, such as where patterns across cases matter more than a deeply reflective understanding of a single pivotal case. While conventional legal scholarship analyzes issues presented in one case or a small group of exceptional or weighty cases, content analysis works by analyzing a larger group of similarly weighted cases to find overall patterns.

Content analysis also assumes an equality among readers of judicial opinions. The traditional legal scholarly enterprise relies, like literary interpretation, on the interpreter's authoritative expertise to select important cases and to draw out noteworthy themes and potential social effects of decisions. Content analysis requires the researcher to explain the selection of cases and themes in enough objective detail to allow others to replicate the steps. This method's persuasiveness depends on the community's ability to reproduce the findings rather than the author's rhetorical power to proclaim them. Content analysis, however, does not displace traditional interpretive legal scholarship. Instead, it offers distinctive insights that complement the types of understanding that only traditional analysis can generate.

In Part III, we propose the best practices for using content analysis in ways that meet both the rigorous standards of social science and the practical needs of legal researchers. This review takes us through the main components of content analysis—selecting cases for study, coding cases to record consistent information about each one, establishing the reliability or replicability of the choices made during the coding, and analyzing data. This article relies heavily on our own content analysis of content analyses. We systematically coded information from one hundred thirty-four projects that perform content analysis of judicial opinions to illustrate numerous points throughout the article.

In the fashion of reporters for the classic Restatement projects of the American Law Institute,⁶ we hope to describe past practices and to point the way to a better future.⁷ The task of a Restatement reporter balances two competing objectives: first, to describe systematically the law in a particular area, and second, to highlight promising developments and encourage more uniform reliance on the most appealing parts of past practice.⁸ We aspire in this

6. In case there is any confusion on this point, this paper is not affiliated with the American Law Institute.

7. Cf. Charles M. Haar, et al., *Computer Power and Legal Reasoning: A Case Study of Judicial Decision Prediction in Zoning Amendment Cases*, 2 AM. B. FOUND. RES. J. 651, 742 (1977) (referring to their content analysis as a "computerized restatement" of zoning law).

8. See AMERICAN LAW INSTITUTE, CAPTURING THE VOICE OF THE AMERICAN LAW INSTITUTE: A HANDBOOK FOR ALI REPORTERS AND THOSE WHO REVIEW THEIR WORK 5 (2005) (Faced with inappropriate or inconsistent precedent, the Institute Reporter is "expected to propose the better rule and provide the rationale for choosing it. A significant contribution of the Restatements has also been anticipation of the direction in which the law is tending and expression of that development in a manner consistent with previously established principles."); Richard L. Cupp, Jr., *Proximate Cause, The Proposed Basic Principles Restatement, and Products Liability*, 53 S.C. L. REV. 1085, 1088 (2001) (explaining the role of Restatement Reporter).

article to earn a similar dual role by collecting numerous examples of studies that rely on content analysis and by describing the long-term trends in the use of this methodology. Just as the Restatement imposes order on the case law to highlight the most desirable trends, we hope that our review of content analysis studies makes the most promising examples more influential in the future.

I

THE DEVELOPMENT OF CONTENT ANALYSIS IN LEGAL SCHOLARSHIP

Content analysis developed decades ago as a method to systematically read and analyze texts.⁹ Some of its earliest users worked in sociology and political science, and this method is now used widely in the communications field as well.¹⁰ Researchers can apply content analysis to texts of any kind, including such legal documents as trial court records, statutes and regulations.¹¹ However, we focus here solely on the most common use of the technique in legal scholarship: coding judicial opinions.

A. Early Practitioners of Content Analysis

Political scientist Fred Kort, in a 1957 article, made the first self-conscious use of content analysis to explore written judicial opinions. Kort collected all the U.S. Supreme Court opinions that discussed the constitutional right to legal counsel in criminal cases, a total of twenty-eight cases decided between 1932 and 1956.¹² Exploring the potential for using computers to predict case outcomes, he developed a coding scheme to record and categorize various facts discussed in the opinions. The Court's opinions flagged these facts as relevant to whether due process required a trial court to appoint a defense lawyer, and lawyers believed that this jumble of "special

9. For early discussions of this method, see generally THOMAS F. CARNEY, *CONTENT ANALYSIS: A TECHNIQUE FOR SYSTEMATIC INFERENCE FROM COMMUNICATIONS* (1972); OLE R. HOLSTI, *CONTENT ANALYSIS FOR THE SOCIAL SCIENCES AND HUMANITIES* (1969); William A. Scott, *Reliability of Content Analysis: The Case of Nominal Scale Coding*, 19 *PUBLIC OPINION Q.* 321 (1955).

10. See KRIPPENDORFF, *supra* note 3, at 5-10.

11. *E.g.*, MICHAEL A. REBELL & ARTHUR R. BLOCK, *EDUCATIONAL POLICY MAKING AND THE COURTS: AN EMPIRICAL STUDY OF JUDICIAL ACTIVISM* (Univ. Chicago Press, 1982) (coding pleadings and briefs); David C. Baldus, et al., *Racial Discrimination and the Death Penalty in the Post-Furman Era: An Empirical and Legal Overview, with Recent Findings from Philadelphia*, 83 *CORNELL L. REV.* 1638 (1998) (coding case file documents); Scott F. Norberg, *Consumer Bankruptcy's New Clothes: An Empirical Study of Discharge and Debt Collection in Chapter 13*, 7 *AM. BANKR. INST. L. REV.* 415, 418 (1999) (coding case file documents); Richard A. Posner, *A Theory of Negligence*, 1 *J. LEGAL STUD.* 29 (1972) (coding trial court records); Sudha Xirasagar, et al., *Small Group Health Insurance: Ranking the States on the Depth of Reform*, 19 *J. HEALTH & SOC. POL'Y* 1 (2004) (coding statutes).

12. The starting point of Kort's study was the decision in *Powell v. Alabama*, 287 U.S. 45 (1932), which first cast the presence and quality of defense counsel as a federal constitutional requirement under the due process clause. For a discussion of this now-defunct body of law, see ANTHONY LEWIS, *GIDEON'S TRUMPET* (1964).

circumstances” produced unpredictable results. Kort used the coding results from half of the cases to develop a scoring system that would allow a reader to predict the outcome of similar cases. He then used that scoring system to correctly predict the outcomes of twelve of the fourteen remaining cases in his collection, an 86% accuracy rate.¹³

Within a few years, other political scientists like Stuart Nagel and Sidney Ulmer joined Kort, using similar techniques to code judicial opinions, typically Supreme Court cases, on various subjects.¹⁴ These early practitioners fell into two camps. Some, like Kort, hoped to use content analysis to predict outcomes based on the facts discussed in the opinions. Others questioned the fundamental premise of principled judicial decision making—the notion that judges apply neutral principles of law to the facts of the case—and instead sought to understand the political or personal attitudes that drove judicial behavior.¹⁵ These early efforts reflected the move toward quantitative methods in political science and the social sciences generally.¹⁶ The advent of computers and academic fascination with the potential to develop “expert systems” using “artificial intelligence” also motivated some of the early users of this technique.¹⁷

About the same time, in the late 1950s and early 1960s, a few lawyers and legal scholars spontaneously began to develop a self-taught method that could

13. Fred Kort, *Predicting Supreme Court Decisions Mathematically: A Quantitative Analysis of the "Right to Counsel" Cases*, 51 AM. POL. SCI. REV. 1, 11 (1957). The two cases that Kort's method did not predict fell into an “indeterminate zone” in the scoring system. *Id.*

Kort's effort may have been the first self-conscious use of content analysis, but there were earlier precedents. Every lawyer is familiar with a precursor form of content analysis that has been widely used for well over a century: West Publisher's Key Number System, and Shepard's Citations. Both are efforts to read cases systematically and to record consistent information about the cases by using pre-existing categories. For a history of the Key Numbers system, see WEST, WEST'S ANALYSIS OF AMERICAN LAW (rev. ed. 2005); West, *The Key Number System: Yesterday and Today*, <http://lawschool.westlaw.com/KNumbers/history.asp?mainpage=16&subpage=4> (last visited July 15, 2007). For discussions of Shepard's, see James F. Spriggs II & Thomas G. Hansford, *Measuring Legal Change: The Reliability and Validity of Shepard's Citations*, 53 POL. RES. Q. 327 (2000); LexisNexis, *Research Solutions: Shepard's Citation Service*, <http://www.lexisnexis.com/shepards> (last visited July 15, 2007).

14. See Glendon Schubert, *Jackson's Judicial Philosophy: An Exploration in Value Analysis*, 59 AM. POL. SCI. REV. 940 (1965); Glendon Schubert, *The 1960 Term of the Supreme Court: A Psychological Analysis*, 56 AM. POL. SCI. REV. 90 (1962); Symposium: *Social Science Approaches to the Judicial Process*, 79 HARV. L. REV. 1551 (1966); S. Sidney Ulmer, *The Analysis of Behavior Patterns on the United States Supreme Court*, 22 J. POL. 629 (1960); S. Sidney Ulmer, *Quantitative Analysis of Judicial Processes: Some Practical and Theoretical Applications*, 28 J. LAW & CONTEMP. PROBS. 164 (1963); S. Sidney Ulmer, *Scaling Judicial Cases: A Methodological Note*, 4 AM. BEHAV. SCIENTIST 31 (1961).

15. See Werner F. Grunbaum & Albert Newhouse, *Quantitative Analysis of Judicial Decisions: Some Problems in Prediction*, 3 HOUS. L. REV. 201 (1965); Stuart Nagel, *Applying Correlation Analysis to Case Prediction*, 42 TEX. L. REV. 1006 (1964); Stuart Nagel, *Predicting Court Cases Quantitatively*, 63 MICH. L. REV. 1411 (1965); Stuart Nagel, *Using Simple Calculations To Predict Judicial Decisions*, 7 PRAC. LAW. 68 (1961).

16. KRIPPENDORFF, *supra* note 3.

17. See Part II.B.4 *infra* (discussing origins and uses of term “jurimetrics”).

be labeled content analysis. These methods were home-grown with no effort to draw on established social science techniques for content analysis. For example, in a 1956 analysis of patent invalidity decisions,¹⁸ attorney P.J. Federico moved beyond traditional legal commentary on case law into a more systematic reading of the texts. Instead of searching for noteworthy themes in “leading” cases, Federico read a more comprehensive set of cases—the fifty most recent examples of Courts of Appeal invalidating a patent—and recorded consistent information about each of them, including the stated grounds for invalidity. Others did similar work in medical malpractice, coding and tabulating basic information about almost two thousand published opinions over a span of one hundred and sixty years.¹⁹

B. Expansion of Content Analysis into Different Subject Areas

During the 1960s and 1970s, the content analysis of judicial opinions expanded into several different subject areas. Legal scholars systematically selected and coded cases in the fields of labor law and zoning.²⁰ Courts’ use of non-legal sources and various other judicial methods also occupied some early case coders.²¹ A few famous scholars took up this method early on: Karl Llewellyn created a version of content analysis to study judicial rhetoric and decision making,²² while Richard Posner’s seminal study of negligence law relied on the coding of 1,528 cases involving accidents between 1875 and 1905.²³

The collections of judicial opinions grew by the late 1970s. In contrast with the early political science studies that focused on U.S. Supreme Court cases dealing with a particular subject, researchers branched out into the state courts and the lower federal courts, sometimes coding hundreds or even thousands of cases.²⁴ The growth in the number of coded cases resulted in part

18. See P.J. Federico, *Adjudicated Patents, 1948-54*, 38 J. PAT. OFF. SOC’Y 233 (1956).

19. Andrew A. Sandor, *The History of Professional Liability Suits in the United States*, 163 JAMA 459 (1957); C. Joseph Stetler, *The History of Reported Medical Professional Liability Cases*, 30 TEMPLE L. Q. 366 (1957).

20. See Grunbaum & Newhouse, *supra* note 15; Haar et al., *supra* note 7.

21. E.g., Neil N. Bernstein, *The Supreme Court and Secondary Source Material: 1965 Term*, 57 GEO. L.J. 55 (1968); Richard A. Daynard, *The Use of Social Policy in Judicial Decision-Making*, 56 CORNELL L. REV. 919 (1971).

22. KARL N. LLEWELLYN, *THE COMMON LAW TRADITION: DECIDING APPEALS* 102-103 (1960)

23. See Posner, *supra* note 11, at 29.

24. E.g., Marc A. Franklin, *Winners and Losers and Why: A Study of Defamation Litigation*, 5 AM. B. FOUND. RES. J. 455 (1980) (680 cases); Robert A. Kagan, et al., *The Business of State Supreme Courts, 1870-1970*, 30 STAN. L. REV. 121 (1977) (5904 cases). The longest coding project is the United States Supreme Court Judicial Database, which political scientist Harold Spaeth began in the mid-1980s and which continues today. See Harold J. Spaeth, *Outputs of the Court*, in *STUDIES IN U.S. SUPREME COURT BEHAVIOR* 291-99 (Harold J. Spaeth & Saul Brenner, eds.) (1990). See also HAROLD J. SPAETH, *THE S. SIDNEY ULMER PROJECT: U.S. SUPREME COURT DATABASES*, <http://www.as.uky.edu/polisci/ulmerproject/sctdata.htm> (last visited

from the computer databases that became available during the late 1970s. Beginning in the 1980s, a number of scholars adopted content analysis to take advantage of Lexis and Westlaw²⁵ along with software that could perform the tedious mathematical calculations involved in statistical analysis.

Legal scholars came to content analysis in greater numbers in the 1990s, and its accelerated use continues.²⁶ According to Table 1, until the 1990s no more than one new project per year was published. The rate of publication jumped to 5.7 per year in the 1990s and stands at eight per year through the middle of 2006.

Looking back, we can now see that legal content analysts have assembled an impressive collection of studies over the years. Though some projects have been published in obscure places, others have also been published in the very best law journals.²⁷ Readers of the literature might not appreciate just how

July 15, 2007) It now consists of almost 9000 decisions—every one since 1953. The largest, by number of decisions, are the State Supreme Court Database and the Court of Appeals Database. Each is a random sample of over 20,000 decisions. Further information about this database is available at <http://www.as.uky.edu/polisci/ulmerproject/apctdata.htm>. See generally, Paul Brace & Melinda Gann Hall, *Comparing Courts Using the American States*, 83 JUDICATURE 250 (2000), available at <http://www.ruf.rice.edu/~pbrace/statecourt/>; Tracey E. George & Reginald Sheehan, *Circuit Breaker: Deciphering Courts of Appeals Decisions Using the U.S. Courts of Appeals Data Base*, 83 JUDICATURE 240 (2000). Although not meeting our definition of content analysis, it is worth noting Thomas Smith's extraordinary database that consists of citation patterns from nearly four million cases. Thomas A. Smith, *The Web of Law*, San Diego Legal Studies Research Paper No. 06-11 (2005), available at http://papers.ssrn.com/abstract_id=642863).

25. See, e.g., Peggy C. Davis, *"There is a Book Out ...": An Analysis of Judicial Absorption of Legislative Facts*, 100 HARV. L. REV. 1539 (1987); Peter Schuck & Donald Elliott, *Studying Administrative Law: A Methodology for, and Report on, New Empirical Research*, 42 ADMIN. L. REV. 519 (1990); Vicki Schultz, *Telling Stories About Women and Work: Judicial Interpretations of Sex Segregation in the Workplace in Title VII Cases Raising the Lack of Interest Argument*, 103 HARV. L. REV. 1749 (1990).

26. Similarly, in the communications field, where content analysis is used extensively, one leading journal experienced a six-fold increase over twenty-four years in the proportion of its articles using content analysis. KIMBERLY A. NEUENDORF, *THE CONTENT ANALYSIS GUIDEBOOK* 27 (2002) (attributing the large increase in content analysis in part to the fact that so many good archives of material are now accessible by computer, especially through the internet).

27. With apologies to those we omit, see, for example, Davis, *supra* note 25 (Harvard Law Review); Schultz, *supra* note 25 (Harvard Law Review); Frank B. Cross & Emerson H. Tiller, *Judicial Partisanship and Obedience to Legal Doctrine: Whistleblowing on the Federal Courts of Appeals*, 107 YALE L.J. 2155 (1998); William Eskridge, *Overriding Supreme Court Statutory Interpretation Decisions*, 101 YALE L.J. 331 (1991); Thomas W. Merrill, *Judicial Deference to Executive Precedent*, 101 YALE L.J. 969 (1992); Joseph A. Grundfest & A.C. Pritchard, *Statutes With Multiple Personality Disorders: The Value of Ambiguity in Statutory Design and Interpretation*, 54 STAN. L. REV. 627 (2002); Kagan, *supra* note 24 (Stanford Law Review); Jane S. Schacter, *The Confounding Common Law Originalism in Recent Supreme Court Statutory Interpretation: Implications for the Legislative History Debate and Beyond*, 51 STAN. L. REV. 1 (1998); Peter H. Schuck & Theodore Hsien Wang, *Continuity and Change: Patterns of Immigration Litigation in the Courts, 1979-1990*, 45 STAN. L. REV. 115 (1992); Peter J. Hammer & William M. Sage, *Antitrust, Health Care Quality, and the Courts*, 102 COLUM. L. REV. 545 (2002); Robert A. Hillman, *Questioning the "New Consensus" on Promissory Estoppel: An Empirical and Theoretical Study*, 98 COLUM. L. REV. 580 (1998); Robert E. Scott, *A Theory of Self-Enforcing Indefinite Agreements*, 103 COLUM. L. REV. 1641 (2003); Barton Beebe, *An*

widespread the method has become, since most studies only cite to a few examples of this research technique, if they cite to any at all.²⁸ To our knowledge, no one has ever compiled a comprehensive list of these studies. We collected all examples of content analysis studies we could locate and coded them for pertinent features, as described in an online Appendix.²⁹ Table 1

Empirical Study of the Multifactor Tests for Trademark Infringement, 94 CALIF. L. REV. 1581 (2006); Frank B. Cross, *Decisionmaking in the U.S. Circuit Courts of Appeal*, 91 CALIF. L. REV. 1457 (2003); Richard L. Revesz, *Congressional Influence on Judicial Behavior? An Empirical Examination of Challenges to Agency Action in the D.C. Circuit*, 76 N.Y.U. L. REV. 1100 (2001); Gregory C. Sisk, et al., *Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning*, 73 N.Y.U. L. REV. 1377 (1998); Jason S. Johnston, *The Statute of Frauds and Business Norms: A Testable Game-Theoretic Model*, 144 U. PA. L. REV. 1859 (1996); R. Polk Wagner & Lee Petherbridge, *Is the Federal Circuit Succeeding? An Empirical Assessment of Judicial Performance*, 152 U. PA. L. REV. 1105 (2004); Sam Kamin, *Harmless Error and the Rights/Remedies Split*, 88 VA. L. REV. 1 (2002); Kimberly D. Krawiec & Kathryn Zeiler, *Common-Law Disclosure Duties and the Sin of Omission: Testing the Meta-Theories*, 91 VA. L. REV. 1795 (2005); Richard L. Revesz, *Environmental Regulation, Ideology, and the D.C. Circuit*, 83 VA. L. REV. 1717 (1997); Wendy Parker, *The Future of School Desegregation*, 94 NW. U. L. REV. 1157 (2000).

28. See, for example, our own two coding projects: Mark A. Hall, et al., *Judicial Protection of Managed Care Consumers: An Empirical Study of Insurance Coverage Disputes*, 26 SETON HALL L. REV. 1055 (1996); Ronald F. Wright & Paul Huck, *Counting Cases About Milk, Our "Most Nearly Perfect" Food, 1860-1940*, 36 LAW & SOC'Y REV. 51 (2002).

29. Mark A. Hall & Ronald F. Wright, *Appendix to Systematic Content Analysis of Judicial Opinions*, Wake Forest Univ. Legal Studies Paper No. 913336 (2006), available at <http://papers.ssrn.com/abstract=913336>. We attempted to identify all empirical projects with publicly available findings as of June 30, 2006 involving systematic content analysis of legal decisions by courts or administrative adjudicative bodies. Systematic coding consists of recording the content of these decisions in a fashion that was analyzed quantitatively using at least simple percentages or proportions, rather than just reporting the raw number of cases in various categories. We included only projects that recorded information beyond merely the subject matter and outcome of the case and the parties' identities. In other words, we included projects that coded for legal, factual, analytic, or linguistic elements of legal decisions that could be gleaned only by a close reading of the opinions, rather than, for instance, information available in a digest or abstract of the decision. To illustrate, we excluded a study of patent litigation that coded for patent subject matter, case outcome, and procedural posture, but we included a similar study that also coded the grounds for finding patent invalidity, grounds such as obviousness or prior art.

As our unit of analysis, we use a case coding project, which is a set of cases gathered and coded for a specified purpose, realizing that a single project can result in multiple publications. We located publications describing these case-coding projects primarily by searching the Journals and Law Reviews (JLR) database in Westlaw, using search terms such as <"content analy!"> or <"empirical w/10 cod!">. We also reviewed the Empirical Legal Studies blog (www.elsblog.org) for references to relevant publications and projects, and we queried or reviewed the bibliographies of researchers who use these coding techniques regularly. We reviewed all results from our Westlaw searches back through 1998, and other relevant publications, looking both for whether the publication met our inclusion criteria, and whether it cited older relevant publications, which we also reviewed for inclusion and citation to other potentially relevant publications. We also looked prospectively to locate later publications that cited these earlier works by applying Westlaw's "citing references" function in its KeyCite tool to some of the leading case-coding projects in the legal literature that we felt were more likely to be cited by other researchers. We reviewed these additional instances for inclusion and for any relevant citations to other case-coding projects. We continued this "snowball" process through multiple iterations until we failed to find additional publications cited by or citing to projects that met our criteria. We followed this snowball

summarizes the trends over the years, starting in 1956.

Table 1: Trends over Time in Content Analysis of Judicial Opinions

| | 1956- 1969 | 1970- 1979 | 1980- 1989 | 1990- 1999 | 2000- June 2006 | All Years |
|--|---------------|---------------|---------------|---------------|-----------------------|--------------|
| Total Projects Coded | 10 | 5 | 10 | 57 | 52 | 134 |
| Projects Published per year | 0.7 | 0.5 | 1.0 | 5.7 | 8 | 2.7 |
| First Author Based in Law (%) | 8 (80) | 3 (60) | 7 (70) | 35 (61) | 31 (60) | 84 (63) |
| Any Author with Ph.D. (%) | 5 (50) | 4 (80) | 7 (70) | 41 (72) | 27 (52) | 84 (63) |
| Citation to Social Science Methodology (%) | 4 (40) | 5 (100) | 7 (70) | 28 (49) | 22 (42) | 66 (49) |
| Median (maximum) Cases per Project | 137 (1936) | 300 (5904) | 190 (630) | 170 (8895) | 399 (22,000) | 252 |

The number of content analysis projects we located is even more impressive considering that this collection omits many examples of case-coding empirical work that did not meet our high threshold for true content analysis. We excluded projects using a method we call “docket analysis,” which codes only for information about cases—such as subject matter, parties, and basic outcomes—that could be obtained from docket sheets or brief abstracts.³⁰ We also excluded the numerous studies of courts’ citation practices and patterns.³¹ Instead, our literature review is based on case coding that reaches the substance

technique into whatever literatures it led, including political science, economics, and sociology. As a result, we are reasonably confident that we have identified the vast majority of relevant publications. We systematically coded the content of these publications or research reports for the factors shown in Tables 1 and 2. One of us coded each publication following a set of coding instructions we devised at the outset, but which we applied independently. To test the reliability of our own coding, we each coded a random selection of sixteen of the publications that the other coded for quantifiable factors. The resulting *kappa* statistics are generally acceptable within various social science literatures, except for our coding of the replicability of case selection. In addition to this quantifiable coding, we noted the following elements qualitatively: (1) the project’s primary research purposes and findings; (2) the project’s principal areas of law or legal scholarship; (3) which cases were selected; (4) the most subjective factors that were coded (5) problems encountered in the coding that the authors specially noted; and (6) other notable aspects of the project’s methods or analysis.

30. See, e.g., Kevin M. Clermont & Theodore Eisenberg, *Plaintiphobia in the Appellate Courts: Civil Rights Really Do Differ from Negotiable Instruments*, 2002 ILL. L. REV. 947; Ronald F. Wright, *Trial Distortion and the End of Innocence in Federal Criminal Justice*, 154 U. PA. L. REV. 79 (2005).

31. See note 72 *infra*.

of judicial reasoning as expressed through the legal and factual content of written opinions, essentially the same material used in traditional interpretive legal methods.

Content analysis has proven useful for studying a broad range of legal subject areas. Our collection includes at least seven projects (5% of the total)³² in areas as far-ranging as administrative law,³³ constitutional law,³⁴ corporate and securities law,³⁵ criminal law and procedure,³⁶ contracts,³⁷ employment discrimination,³⁸ health law,³⁹ and torts.⁴⁰ A large number of the studies explicitly focus on questions of legal methods, judicial decision making, and statutory interpretation.⁴¹ The collection is surprisingly thin in areas such as property, international law, and tax.

The spread of content analysis to study judicial opinions also cuts across

32. In addition, five studies focus on patent litigation. *See, e.g.*, GLORIA K. KOENIG, *PATENT INVALIDITY: A STATISTICAL AND SUBSTANTIVE ANALYSIS* (rev. ed. 1980); Wagner & Petherbridge, *supra* note 27. Four studies address family law. *See, e.g.*, Davis, *supra* note 25; Kathryn L. Mercer, *A Content Analysis of Judicial Decision-Making—How Judges Use the Primary Caretaker Standard to Make a Custody Determination*, 5 WM. & MARY J. WOMEN & L. 1 (1998).

33. *See, e.g.*, Thomas W. Merrill, *Judicial Deference to Executive Precedent*, 101 YALE L.J. 969 (1992); Joseph L. Smith & Emerson H. Tiller, *The Strategy of Judging: Evidence from Administrative Law*, 31 J. LEGAL STUD. 61 (2002).

34. *See, e.g.*, Timothy M. Hagle, *But Do They Have to See It to Know It? The Supreme Court's Obscenity and Pornography Decisions*, 44 WESTERN POL. Q. 1039 (1991); Adam Winkler, *Fatal in Theory and Strict in Fact: An Empirical Analysis of Strict Scrutiny in the Federal Courts*, 59 VAND. L. REV. 793 (2006); *infra* note 111.

35. *See, e.g.*, Theresa A. Gabaldon, *A Sense of a Security: An Empirical Study*, 25 J. CORP. L. 307 (2000); John A. Swain & Edwin E. Aguilar, *Piercing the Veil to Assert Personal Jurisdiction Over Corporate Affiliates: An Empirical Study of the Cannon Doctrine*, 84 B.U. L. REV. 445 (2004); note 112 *infra*.

36. *See, e.g.*, Herbert M. Kritzer & Mark J. Richards, *The Influence of Law in the Supreme Court's Search-and-Seizure Jurisprudence*, 33 AMER. POL. RES. 33 (2005); James S. Liebman, et al., *Capital Attrition: Error Rates in Capital Cases, 1973-1995*, 78 TEX. L. REV. 1839 (2000).

37. *See, e.g.*, Robert A. Hillman, *Questioning the "New Consensus" on Promissory Estoppel: An Empirical and Theoretical Study*, 98 COLUM. L. REV. 580 (1998); Jason S. Johnston, *The Statute of Frauds and Business Norms: A Testable Game-Theoretic Model*, 144 U. PA. L. REV. 1859 (1996).

38. *See, e.g.*, James J. Brudney, et al., *Judicial Hostility Toward Labor Unions? Applying the Social Background Model to a Celebrated Concern*, 60 OHIO ST. L.J. 1675 (1999); Vicki Schultz & Stephen Petterson, *Race, Gender, Work and Choice: An Empirical Study of the Lack of Interest Defense in Title VII Cases Challenging Job Segregation*, 59 U. CHI. L. REV. 1073 (1992); note 111 *infra*.

39. *See, e.g.*, Peter D. Jacobson et al., *The Role of the Courts in Shaping Health Policy: An Empirical Analysis*, 29 J.L. MED. & ETHICS 278 (2001); Edward Alan Miller, *Federal Oversight, State Policy Making, and the Courts: An Empirical Analysis of Nursing Facility Litigation Under the Boren Amendment*, 3 J. EMPIRICAL LEGAL STUD. 145, 145-173 (2006); note 113 *infra*.

40. *See, e.g.*, Marc A. Franklin, *Winners and Losers and Why: A Study of Defamation Litigation*, 5 AM. BAR FOUND. RES. J. 455 (1980); Posner, *supra* note 11.

41. *See, e.g.*, Daniel M. Schneider, *Empirical Research on Judicial Reasoning: Statutory Interpretation in Federal Tax Cases*, 31 N.M. L. REV. 325 (2001); Peter Siegelman & John J. Donohue III, *Studying the Iceberg from Its Tip: A Comparison of Published and Unpublished Employment Discrimination Cases*, 24 LAW & SOC'Y REV. 1133 (1990).

disciplinary boundaries. Both social scientists and lawyers use the method to analyze judicial opinions. Among legal researchers, however, about one-third have a Ph.D. and are likely to have been trained in social science methods. Thorough content analysis is demanding, and those with social scientific training are likely to be both better trained and more motivated to take on this painstaking method. However, it is not necessary to practice this method only at its highest level. Instead, we will show that even largely innumerate scholars can find value in using content analysis.

The payoff for scholars who use content analysis is substantial.⁴² Using the number of citations as a crude but widely accepted proxy for the influence of a scholarly work, content analysis projects published during the 1990s generated an average of seventy-seven citations per article, with a mean of thirty citations per project. Considering projects from all decades, 87% received at least one citation and 71% generated at least five citations. These citation patterns compare favorably to the general trends in legal scholarship. According to Thomas Smith's ongoing research on the citation patterns in legal scholarship, 40% of articles receive no citations at all.⁴³ Therefore, content analysis projects appear somewhat more likely to generate discussion and citation than law review articles more generally.⁴⁴

It is striking how often legal researchers employ this method without citing to any methodology literature, or only citing to examples from legal literature. As Table 1 shows, only about one-half of case-coding researchers cite to social science methodology sources, and this happens even less among legal scholars.⁴⁵ In project after project, legal researchers reinvent this methodological wheel on their own.⁴⁶ The two of us, for instance, each learned how to do content analysis on the fly, feeling at first as if we each discovered something new until we learned that we had each done the same thing

42. See note 27, *supra*.

43. Smith, *supra* note 24. See also Paul L. Caron, *The Long Tail of Legal Scholarship*, 116 YALE L.J. Pocket Part 38 (2006), available at <http://www.thepocketpart.org/2006/09/06/caron.html>, (last visited July 15, 2007).

44. Our analysis is based on a random sample of forty-five projects, one-third of the total, covered by the LEXIS-NEXIS database. For each project, we determined the number of citations to the first published description of the project found in the LEXIS-NEXIS database, both to secondary literature and to other sources covered in the LEXIS-NEXIS database. We recognize that our pool suffers from a selection bias for purposes of this comparison. As discussed in note 29, one method we used to identify some of the projects was the fact that a study was cited in another article that used content analysis. Thus, we are somewhat less likely to find articles that received no citations at all in our pool.

45. Social science methodology was cited in only 26% (13/50) of projects with lawyers as first authors and where no authors had Ph.D.s (although another 20% (10) of these projects did cite to methodology sources in the legal literature). However, over one-half of legal scholars without Ph.D. co-authors cited to no methodology literature of any kind. This compares with only 23% (19/84) of projects with Ph.D. authors which cited to no methodology literature, and 63% (63/84) of these which cited to social science methodology.

46. *Accord* Hammer & Sage, *supra* note 27 at 560 (noting the "tendency for each new enterprise to invent its own wheel, often in a fairly ad hoc manner").

independently. We see now that many of our colleagues share the same sense of having found their own way.

Legal scholars using content analysis often describe their steps at excruciating length in a way that social scientists would view as absurd. Articles commonly approach or exceed one hundred pages, laboriously detailing how the author devised techniques and resolved quandaries.⁴⁷ Others note that this has resulted in a “confused area of scholarship, with few clear norms or standards to guide researchers.”⁴⁸ This profusion of do-it-yourself research techniques points to a need for methodological tenets that legal scholars can share, learn, refine, and cite briefly.

In retrospect, it is not surprising that legal scholars transitioned to content analysis from the more familiar interpretive analysis of law. These distinct approaches to reading cases hold much in common. In the classic interpretive method of analysis, the legal scholar typically comments on the significance of multiple opinions, because as Karl Llewellyn’s famous introductory text for law students states, “no case can have a meaning by itself!”⁴⁹ The interpretive legal scholar reads opinions closely, looking for common themes running through several opinions.

The writers of legal treatises take the interpretive legal method furthest down the path to content analysis. Treatise writers must read large groups of cases and generally must convince readers that the cases they discuss are representative of the cases that readers will find in their home jurisdictions. It is only a few short steps from treatise researcher to content analyst. Thus, it is easy to see now how legal scholars slipped into content analysis, even if they did not call the method by its name, as a natural extension of their intellectual curiosity and search for a more empirically grounded understanding of case law.

Several leading examples of content analysis come from legal scholars who once approached issues more conventionally but who, out of a sense of “intellectual honesty,”⁵⁰ felt the need for a surer epistemological basis to support their claims or to question others. James Henderson and Ted Eisenberg, for instance, found their way to this method in the late 1980s while doing

47. See, e.g., Haar et al., *supra* note 7; Andrew P. Morriss, *Developing a Framework for Empirical Research on the Common Law: General Principles and Case Studies of the Decline of Employment at Will*, 45 CASE W. RES. L. REV. 999 (1995); Schuck & Elliott, *supra* note 25; Sisk et al., *supra* note 27; Gregory C. Sisk, Michael Heise, & Andrew P. Morriss, *Searching for the Soul of Judicial Decisionmaking: An Empirical Study of Religious Freedom Decisions*, 65 OHIO ST. L. J. 491 (2004).

48. Hammer & Sage, *supra* note 27 at 560.

49. KARL N. LLEWELLYN, *THE BRAMBLE BUSH: ON OUR LAW AND ITS STUDY* 49 (1960) (“What counts, what gives you leads, what gives you sureness, *that is the background of the other cases* in relation to which you must read the one.”) (emphasis in original).

50. Laura E. Little, *Hiding with Words: Obfuscation, Avoidance, and Federal Jurisdiction Opinions*, 46 UCLA L. REV. 75 (1998).

conventional research to update their treatise.⁵¹ Noticing an apparent shift in courts' attitudes about products liability, they felt the need to document the trend more thoroughly by coding and statistically analyzing all relevant cases over six years. A decade later, Robert Hillman found himself unconvinced by new scholarship claiming that promissory estoppel was overtaking consideration as the basis for enforcing promises, even in the absence of detrimental reliance. He therefore undertook to code all of the relevant cases for two years to find out if this was really so (it was not).⁵²

Although the treatise writer and the content analyst have much in common, each offers different insights. When Dean Prosser read cases for possible discussion in his Torts treatise,⁵³ he was auditioning a crowd of singers to find the best soloists. His objective was to select particular cases that eloquently stated rules of law or illustrated a trend. Content analysts, on the other hand, do not audition soloists. Instead, they assemble a chorus, listening to the sound that the cases make together. This distinction between the collective and individual insights drawn from judicial opinions is the starting point for the functional differences between content analysis and traditional literary legal analysis, a distinction we now explore in more detail.

II

THE EPISTEMOLOGY OF CONTENT ANALYSIS

The epistemological roots of content analysis lie in Legal Realism, the school of jurisprudence that rejects Legal Formalism's search for independent doctrines of law that constrain legal actors.⁵⁴ Holmes famously proclaimed that "prophecies of what the courts will do in fact, and nothing more pretentious, are what I mean by the law."⁵⁵ While Holmes himself was not an empiricist, this credo's call to empirical methods is obvious. At the outset, this article quotes Herman Oliphant, a leader among the first Legal Realists, exhorting law professors to employ social science tools in the study of what courts do. Most of the original Legal Realists, however, never heeded this call to empiricism.⁵⁶ They remained occupied with jurisprudential attacks on formalism, and during

51. James A. Henderson & Theodore Eisenberg, *The Quiet Revolution in Products Liability: An Empirical Study of Legal Change*, 37 UCLA L. REV. 479 (1990).

52. Hillman, *supra* note 27.

53. WILLIAM L. PROSSER, TORTS (3d ed. 1964).

54. See generally Brian Leiter, *Rethinking Legal Realism: Toward a Naturalized Jurisprudence*, 76 TEXAS L. REV. 267 (1997).

55. Oliver Wendell Holmes, Jr., *The Path of the Law*, 10 HARV. L. REV. 457, 461 (1897).

56. See Leiter, *supra* note 54, at 311-12; Stewart Macaulay, *The New Versus the Old Legal Realism: "Things Ain't What They Used to Be,"* 2005 WIS. L. REV. 365, 374-77; John Henry Schlegel, *American Legal Realism and Empirical Social Science: From the Yale Experience*, 28 BUFF. L. REV. 459, 513-19 (1979). A recent effort called New Legal Realism, based at the University of Wisconsin, attempts to re-energize this field of scholarship with a rich array of interdisciplinary social science methods. For a discussion of their work, see Symposium, *New Legal Realism Symposium*, 2005 WIS. L. REV. 335 (2005).

the Great Depression they devoted time to their social reform agenda. The tools available to legal scholars also did not fit the Legal Realist agenda as well as they might have hoped because social science methods were not as fully developed then.

Perhaps another reason Legal Realists failed to employ empirical methods is that they never needed to do much of this hard work. Legal Realists handily won their jurisprudential battle against Legal Formalists, so now it is sometimes said in the legal academy that "we are all legal realists."⁵⁷ Legal Realists' initial claims about the importance of studying judicial behavior, once revolutionary, are now so widely accepted that they seem banal. Accordingly, making empirical claims about the actual thinking and behavior of judges in deciding cases, as partially revealed in their opinions, is now completely commonplace when analyzing case law. Using empirical methodology to support these claims means that scholars are only beginning to catch up with jurisprudential movements that began a century ago.

Content analysis holds value not only for conventional doctrinal analysis, but also for more theoretically influenced work in major branches of jurisprudence, such as economic analysis or critical theory. Writers in each of these scholarly camps frequently claim, for instance, that judges and the law respond predictably to various social, political and market conditions—empirical claims that researchers can systematically test.⁵⁸ Content analysis allows scholars to verify or refute the empirical claims about case law that are implicit or explicit in all branches of legal scholarship. Some schools of jurisprudence emphasize the bare outcomes of cases in relation to their raw facts, while others emphasize how judges explain their decisions. Although these fundamental differences hugely affect how researchers might employ content analysis, the basic method is adaptable to any branch of legal theory that systematically studies judicial opinions. Content analysis is not tied to any particular jurisprudential school, other than to positivism in the broadest sense.⁵⁹

57. Leiter, *supra* note 54.

58. From a law and economics perspective, see, for example, Johnston, *supra* note 27; Fred S. McChesney, *Tortious Interference with Contract versus "Efficient" Breach: Theory and Empirical Evidence*, 28 J. LEG. STUD. 131 (1999); Daniel T. Ostas, *Predicting Unconscionability Decisions: An Economic Model and an Empirical Test*, 29 AM. BUS. L. J. 535 (1991). From perspectives of critical theory, psychology, or social theory, see, for example, Eskridge, *supra* note 25; Little, *supra* note 50; Willy E. Rice, *Race, Gender "Redlining," and the Discriminatory Access to Loans, Credit, and Insurance: An Historical and Empirical Analysis of Consumers Who Sued Lenders and Insurers in Federal and State Courts, 1950-1995*, 33 SAN DIEGO L. REV. 583 (1996); Kimberly Richman, *Lovers, Legal Strangers, and Parents: Negotiating Parental and Sexual Identity in Family Law*, 36 LAW & SOCIETY REV. 285 (2002).

59. By positivism, we mean *logical* positivism, the basic philosophical view of the world underlying modern science, as opposed to *legal* positivism, which examines factors that determine the legitimacy of law. Acknowledging the usefulness of content analysis as a general empirical method does not necessarily entail any claim about the jurisprudential importance of the particular aspects of judicial opinions that a researcher might choose to study, only that some aspect of

Content analysis does not hold value, however, for every scholarly project involving judicial opinions. Its strength is to provide an objective understanding of a large number of decisions where each decision has roughly the same value. Lawyers and legal scholars, though, often do not make empirical claims about a body of case law that they intend to be scientifically verifiable. Instead, they mean to analyze particular decisions in a subjective, interpretative, or advocacy mode akin to the way a literary critic interprets poetry or the way a social reformer critiques a political speech. These rhetorical goals are not advanced meaningfully by systematic, objective verification of claims about the content of a collection of judicial opinions. Although legal writing in this mode may contain assertions about how judges think or act, these are “empirical” only in a casual way.⁶⁰ These legal analysts report and interpret what they see in key cases, without necessarily claiming that their observations are prevalent in the law or typical of other cases. For this kind of scholarship, systematic content analysis would be either irrelevant or overkill. A simple citation to the relevant sources can usually verify the minimally empirical claims in conventional legal scholarship.

Still, as Kay Levine insists, “is it not our obligation as academics” to ask, “[c]an anyone know the state of the law from reading a handful of select cases?”⁶¹ She explains the “malady” of basing our understanding only on traditional case analysis—because theories based on selective readings “are generated after the fact to fit a small series of cases and rely only on outsider observations of what those cases mean, [they] leave us with little reason for optimism about their predictive abilities for future cases. While [these theories] have the capacity to be thought-provoking, they ultimately prove unsatisfying to scholars seeking a more robust explanation of how the law works.”⁶²

The goal of this Part is to identify when this type of collective, objective understanding is useful—in short, which kinds of legal questions bring out the best in this method? We will map this epistemological landscape taking two routes: first, we dissect content analysis to learn the limits of what each of its particular components can contribute to our understanding of case law, and second, we broadly survey the kinds of questions that content analysis, used at full force, can and cannot answer well.

judicial opinions understood empirically have jurisprudential importance. *See generally* Brian Leiter, *Positivism, Formalism, Realism*, 99 COLUM. L. REV. 1138 (1999) (discussing the connection between realism and positivism).

60. Jack Goldsmith & Adrian Vermeule, *Empirical Methodology and Legal Scholarship*, 69 U. CHI. L. REV. 153, 155-57 (2002).

61. Kay L. Levine, *The Law is Not the Case: Incorporating Empirical Methods into the Culture of Case Analysis*, 17 U. FLA. J.L. & PUB. POL'Y 283, 284 (2006).

62. *Id.* at 300.

A. *The Components of Content Analysis*

There are three distinct components of content analysis: (1) selecting cases; (2) coding cases; and (3) analyzing the case coding, often through statistical methods. Each component contributes something of value, yet at each of these three stages, some legal scholars have raised objections. The critics question whether content analysis adds anything to traditional interpretive methods, or they claim that it obscures important insights or produces false results. Rather than questioning whether a particular study *executes* the method well, they ask whether the game is worth the candle. This ground-level challenge to content analysis calls on us to compare each of the three components of content analysis with the traditional interpretive reading of judicial opinions.

1. *Systematic Case Selection*

At the most basic level, empirically minded legal scholars can be more systematic in selecting cases for analysis. Interpretive legal scholars present the cases that interest them, often with no discussion at all about where they found the cases or why they selected them over other candidates for discussion. The reader depends on the author's judgment about which cases are worth reading, that is, which are the "leading cases" that best illustrate the legal issue in question. Content analysis, in sharp contrast, insists on replicability. Empirical legal scholars can achieve this first principle simply by specifying a reproducible selection of cases. This entails deciding which opinions might answer the research question, articulating the selection criteria, and then reading the opinions.

A prominent historical example is Karl Llewellyn's 1960 book *The Common Law Tradition*,⁶³ which records his observations after systematically reading the "mine run" of thousands of cases randomly selected from courts across the country. Llewellyn's goal was a more accurate understanding of which types of cases courts were hearing and what rhetorical styles courts used to decide them.⁶⁴ Similarly, Landes' and Posner's classic *The Economic Structure of Tort Law* contains multiple examples of systematic case selection, such as reading every tenth case that cites Judge Hand's *T.J. Hooper* opinion on custom, every common law tort case affirming a punitive damages award in the most recent volume of each West regional reporter, or every federal appellate products liability case over three years.⁶⁵ More recently, conventional

63. LLEWELLYN, *supra* note 22.

64. *Id.* at 3-6.

65. WILLIAM M. LANDES & RICHARD A. POSNER, *THE ECONOMIC STRUCTURE OF TORT LAW* 103, 135-37, 184-85, 302 (1987). Also starting in the 1980s, a series of articles in leading law reviews explored developments in the law of promissory estoppel by systematically reading all relevant cases over defined time spans. See Daniel A. Farber & John H. Matheson, *Beyond Promissory Estoppel: Contract Law and the "Invisible Handshake,"* 52 U. CHI. L. REV. 903, 909

non-empirical legal scholars increasingly take advantage of computerized legal databases to specify the universe of cases they research.⁶⁶

We are not calling for scholars to publish each research step they take; law review articles are too long already. But it is good research discipline to define a relevant selection of cases beforehand and to document how they were located,⁶⁷ just as a careful historian might do, even if it is not necessary to include the documentation in every publication. Disciplined research is more likely to uncover all the relevant cases, and it guards against subliminal biases in selecting only cases that prove the author's point.

2. *Systematic Case Coding*

Content analysis requires scholars to read their selected cases more consistently and with greater focus on their announced research questions. Users of this method create a coding scheme and record defined elements of each case. As Harvard Law Professor Charles Haar explains:

Flights of intellectual fancy (or tentative hypotheses) as to which factors persuade particular courts, based on their past decisions, can be tested and evaluated. Meaningful insight into the cases can be confirmed statistically, and areas of uncertainty can be highlighted. Although it is no substitute for legal analysis, the disciplined reading and analysis of the cases required to code them for computer analysis

(1985) (reading every case over a ten year period that cited section 90 of the Restatement of Contracts); Jay M. Feinman, *Promissory Estoppel and Judicial Method*, 97 HARV. L. REV. 678 (1984) (reading cases listed under particular West key numbers); Henderson and Eisenberg, *supra* note 51; Hillman, *supra* note 52.

66. *E.g.*, Sidney A. Shapiro & Richard E. Levy, *Judicial Incentives and Indeterminacy in Substantive Review of Administrative Decisions*, 44 DUKE L.J. 1051, 1067 n. 62 (1995) (reviewing all cases found through a Westlaw search that involved judicial review of administrative decisions over a defined time span). As a rough indication of how often this is done, on May 29, 2006 we found 1484 hits from searching Westlaw's Journal and Law Review (JLR) database for the past five years for articles that contained the term "search!" within three words of "Westlaw" or "Lexis." A casual review of the first twenty of these confirmed that most were instances of authors conducting database searches to verify points made in their articles.

67. For instance, in DOUGLAS LAYCOCK, *THE DEATH OF THE IRREPARABLE INJURY RULE* 23-24 (1991), the author explained that:

[o]ne of my students compiled a list of cases with digest entries under the principal West key numbers reciting the irreparable injury rule. She gathered from that universe the hundred most recent federal and the hundred most recent state cases granting equitable relief, and the hundred most recent federal and the hundred most recent state cases denying equitable relief. Another student updated that effort in 1988. To these lists I added the fruits of screening United States Law Week, the Supreme Court, Federal, Texas, and Northeastern advance sheets, and cases from the academic literature. Then I worked backwards and forwards through citation patterns, examining all the cases we found that granted or denied equitable relief. I took care to sample most United States jurisdictions and to get a reasonable number of cases on all issues that emerged in the sample. I found new issues and whole lines of cases I had not anticipated; I did not limit the search to a predetermined set of issues. This sample of cases does not statistically represent some larger population. But I am confident that the cases are broadly representative, and that there is no other line of cases that would give a different picture.

eliminates casual meandering through factors on a case-by-case basis.⁶⁸

A defined coding scheme focuses attention more methodically on various elements of cases and is a check against looking, consciously or not, for confirmation of predetermined positions. This effort to articulate beforehand the features of a case worth studying also allows researchers to delegate some or all of the reading to assistants. More importantly, coding cases, even for just qualitative description and analysis, strengthens the objectivity and reproducibility of case law interpretation.

This degree of rigor in reading cases can be especially helpful when trying to document the absence of an element that is thought to be present in case law. A prominent example is Douglas Laycock's *The Death of the Irreparable Injury Rule*, which demonstrates, based on reading more than 1400 cases, that irreparable injury is not an important factor when courts grant or deny equitable relief.⁶⁹ "Proving a negative" is far harder than simply pointing to what is present in case law because the nay-saying researcher must demonstrate that she has looked systematically for all likely instances of the missing element.

3. Analyzing Cases Quantitatively

At the next stage of content analysis, scholars can quantitatively tabulate information coded from opinions, enabling them to draw conclusions from the features that they find scattered throughout the cases. By counting rather than just reading cases, legal researchers enter the realm occupied by social scientists.⁷⁰ They describe patterns and associations in opinions with some confidence about the validity and reliability of those observations across the broader universe of sampled cases.

Case counting can focus on any aspect of legal decisions, including their bare outcomes, or which authorities they cite. But our survey focused on examples of content analysis that brought some legal judgment to bear on the judicial opinions analyzed, such as describing the content of the parties' arguments or the judge's reasoning, or studying the influence of legally relevant facts. We collected over one hundred thirty examples of this method in an online Appendix.⁷¹ Using one technique or another, each example analyzes the consistently recorded factual or legal content of judicial opinions. Other efforts to analyze some features of judicial decisions did not make the cut

68. Haar et al., *supra* note 7 at 746.

69. See LAYCOCK, *supra* note 67.

70. Social scientists also use rigorous *qualitative* analysis, such as ethnography, but we found only one example based on case content analysis. See Henry F. Fradella, Adam Fogarty & Lauren O'Neill, *The Impact of Daubert on the Admissibility of Behavioral Science Testimony*, 30 PEPP. L. REV. 403 (2002). Most qualitative analyses based on case coding use the looser, discursive style common to conventional legal scholarship, which is not regarded as scientific.

71. See Hall & Wright, *supra* note 29.

because they did not code a sufficient level of detail from the content of opinions to meet our inclusion criteria.⁷²

As discussed in Part III, the precise methods used to quantify cases are subject to a number of criticisms.⁷³ Here, we consider a more general challenge to any case coding effort. Critics decry the false sense of precision or certainty that attaches to systematic content analysis. Calling this “pseudo-measurement,” Wallace Mendelson forcefully argued almost a half century ago that political scientists who embrace this approach to legal studies

fail to depict even dimly the subtleties of the judicial process. They do not, presumably because they cannot, measure the range of values that play in the jurisprudence of a Holmes, a Brandeis, a Stone, or a Cardozo—to mention a few departed heroes. . . . [T]he judge’s art, when greatly practiced, is far too subtle to be measured by any existing behavioral technique. “The law,” said Holmes, “is the painting of a picture—not the doing of a sum.”⁷⁴

As a vivid example, Mendelson points to the failure of political scientists to detect through the most sophisticated coding techniques the Supreme Court’s subtle signals that it was about to change its attitude to right-to-counsel cases just prior to *Gideon v. Wainwright*. While the existing constitutional doctrine still depended nominally on the presence of “special circumstances,” conventional legal scholars, by tuning in to the hints in the most recent line of cases, were able to predict correctly that the Court would abandon the old rule in favor of a broader right to counsel in felony cases.⁷⁵

The most effective response is to acknowledge the validity of these

72. Some code only for case subject matter and outcome. See, e.g., Ruth Colker, *The Americans with Disabilities Act: A Windfall for Defendants*, 34 HARV. C.R.-C.L. L. REV. 99 (1999); Tracey E. George, *Developing a Positive Theory of Decisionmaking on U.S. Courts of Appeals*, 58 OHIO ST. L.J. 1635 (1998); Thomas J. Miles & Cass R. Sunstein, *Do Judges Make Regulatory Policy? An Empirical Investigation of “Chevron,”* 73 U. CHI. L. REV. 823 (2006); Cass R. Sunstein, David Schkade & Lisa Michelle Ellman, *Ideological Voting on Federal Courts of Appeals: A Preliminary Investigation*, 90 VA. L. REV. 301 (2004). Others simply code for which authorities the opinion cites. See, e.g., Patricia J. Falk, *The Prevalence Of Social Science In Gay Rights Cases: The Synergistic Influences Of Historical Context, Justificatory Citation, And Dissemination Efforts*, 41 WAYNE L. REV. 1 8, n. 25 (1994) (collecting citations to other studies of citation patterns); Lawrence M. Friedman, Robert A. Kagan, Bliss Cartwright, & Stanton Wheeler, *State Supreme Courts: A Century of Style and Citation*, 33 STAN. L. REV. 773 (1981); William M. Landes and Richard A. Posner, *Legal Precedent: Theoretical and Empirical Analysis*, 19 J.L. & ECON. 249, 271 (1976); William H. Manz, *Cardozo’s Use of Authority: An Empirical Study*, 32 CALIF. WEST. L. REV. 31 (1995); Frank B. Cross & Thomas A. Smith, *The Reagan Revolution in the Network of Law* (June 2006), available at <http://ssrn.com/abstract=909217>.

73. See *infra* text accompanying notes 148-243.

74. Wallace Mendelson, *The Neo-Behavioral Approach to the Judicial Process: A Critique*, 57 AM. POL. SCI. REV. 593, 595, 602-03 (1963). See also Howard Gillman, *What’s Law Got to Do With It?*, 26 LAW & SOC. INQUIRY 465 (2001) (criticizing the reductionist aspect of case coding, which tends to neglect matters of importance).

75. Wallace Mendelson, *The Untroubled World of Jurimetrics*, 26 J. POLITICS 914, 921-22 (1964).

charges, as far as they go, but to realize that one form of analysis does not displace the other. Instead, content analysis can complement more interpretive forms of legal scholarship.⁷⁶ Social scientists speak in terms of “triangulating” different methods – that is, exploring whether different approaches offer similar conclusions, each approach rigorous in its own way, but each illuminating different dimensions.⁷⁷ Quantitative description can tell us the *what* of case law; other methods may be better suited to understanding the *why* and *wherefore*. Neither type of scholarship standing alone is as strong as the different types combined.⁷⁸

We agree with Karen Jordan, who argues in defense of one of our own coding efforts that “studies of judicial decisions yield useful, albeit narrow information, that moves us toward a greater understanding of the bigger policy questions” and that helps uncover areas for further research.⁷⁹ These methods also have “considerable power for the discovery of anomalies which may escape the naked eye.”⁸⁰ Lon Fuller long ago noted that a “possible gain from researches of this kind lies in the realm of serendipity. A puzzling correlation that violates normal anticipations may set our minds going along new paths and yield unexpected insights.”⁸¹ Once detected, these previously unnoticed and unexpected features of the law, observed only on the surface, can be explored more deeply through other, richer methods.

Still, it is not the case that content analysis is well suited to most legal research projects. It is best used when each decision should receive equal weight—that is, when it is appropriate to regard the content of opinions as generic data. Coding and counting cases usually assumes that the information

76. See James J. Brudney & Corey Ditslear, *Canons of Construction and the Elusive Quest for Neutral Reasoning*, 58 VAND. L. REV. 1 (2005); Reed Dickerson, *Some Jurisprudential Implications of Electronic Data Processing*, 28 L. & CONTEMP. PROBS. 53, 56; Theodore Eisenberg & Sheri Johnson, *The Effects of Intent: Do We Know How Legal Standards Work?* 76 CORNELL L. REV. 1151 (1991); Peter D. Jacobson et al., *The Role of the Courts in Shaping Health Policy: An Empirical Analysis*, 29 J.L. MED. & ETHICS 278 (2001); Ann Juliano & Stewart J. Schwab, *The Sweep of Sexual Harassment Cases*, 86 CORNELL L. REV. 548, 560 (2001); Fred S. McChesney, *Doctrinal Analysis and Statistical Modeling in Law: The Case of Defective Incorporation*, 71 WASH. U.L.Q. 493, 534 (1993).

77. Levine, *supra* note 61, at 286.

78. Lee Epstein, et al., *Judging Statutes: Thoughts on Statutory Interpretation and Notes for a Project on the Internal Revenue Code*, 13 WASH. U. J.O. & POL'Y 305, 323 (2003) (describing a coding project that aims to “merge and incorporate the positive features of the legal and social science programs, while leaving the problematic aspects of each method behind”); Haar et al., *supra* note 7, at 656 (“the knowledge gained by combining legal and statistical analysis allows the analyst to reach conclusions that could not be justified on the basis of either methodology alone”).

79. Karen A. Jordan, *Empirical Studies of Judicial Decisions Serve an Important Role in the Cumulative Process of Policy Making*, 31 IND. L. REV. 81, 88 (1998).

80. Alan L. Tyree, *Fact Content Analysis of Case Law: Methods and Limitations*, 22 JURIMETRICS J. 1, 23 (1981).

81. Lon L. Fuller, *An Afterward: Science and the Judicial Process*, 79 HARV. L. REV. 1604, 1622 (1965-66).

from one opinion is potentially as relevant as that from any other opinion.⁸² Because content analysis tends to regard all cases, judges, courts, and jurisdictions the same, it should be used only with great caution when any of these have a great deal more status or influence than the others for the question addressed.⁸³ Decisions often should carry unequal weight because precedent and persuasiveness depend on various qualitative judgments about the reasons given or the source of the decision.⁸⁴ In short, “[t]he legal and cultural salience of *Roe v. Wade* far outruns its statistical significance.”⁸⁵

Taking these limitations into account, scholars have found it especially useful to code and count cases in studies that debunk conventional legal wisdom. Sometimes authors do this to show they were wrong in their own published preconceptions that were based on conventional scholarship.⁸⁶ Content analysis works well in this setting because it highlights the methodological weakness of traditional legal analysis; that is, this method can simultaneously demonstrate the error of a conventional view and explain how it became the convention.⁸⁷ For instance James Henderson and Theodore Eisenberg documented a sharp shift of judicial attitudes in the 1980s regarding tort liability by showing that products liability litigation was not increasing rapidly as critics often claimed.⁸⁸

Conventional interpretive analysis is sometimes wrong about what appears in the case law or what affects case outcomes. Rather than engaging in a shouting match over whether selected examples amount to a real trend,⁸⁹ a researcher can turn to systematic, objective coding that creates replicable

82. Hammer & Sage, *supra* note 27, at 561; Kagan et al., *supra* note 24, at 125 n.10. It is possible, however, to combine quantitative content analysis with qualitative assessments of the strength or perceived impact of different decisions. See, e.g., Brudney & Ditslear, *supra* note 76; Jacobson et al., *supra* note 76.

83. See *infra* text accompanying notes 242-243 (discussing prospects and difficulties of weighted analysis).

84. Thomas Smith reports, based on analysis of nearly four million cases, that “precedential authority is concentrated in a small number of cases” and that the “vast majority of cases are rarely or never cited.” Smith, *supra* note 24.

85. Goldsmith & Vermeule, *supra* note 60, at 160.

86. See Hall et al., *supra* note 28; Schuck & Wang, *supra* note 27.

87. See, e.g., John R. Allison & Mark A. Lemley, *Empirical Evidence on the Validity of Litigated Patents*, 26 AIPLA Q. J. 185-276 (1998); Beebe, *supra* note 27; Merrill, *supra* note 27; Schuck & Elliot, *supra* note 25; Michael E. Solimine, *The Quiet Revolution in Personal Jurisdiction*, 73 TUL. L. REV. 1 (1998).

88. Henderson & Eisenberg, *supra* note 51.

89. See, e.g., Michael K. Curtis, *The Bill of Rights as a Limitation on State Authority: A Reply to Professor Berger*, 16 WAKE FOREST L. REV. 45 (1980); Raoul Berger, *Incorporation of the Bill of Rights in the Fourteenth Amendment: A Nine-Lived Cat*, 42 OHIO ST. L.J. 435, 463 (1981); Michael K. Curtis, *The Fourteenth Amendment and the Bill of Rights*, 14 CONN. L. REV. 237 (1982); Raoul Berger, *Incorporation of the Bill of Rights: A Reply to Michael Curtis' Response*, 44 OHIO ST. L.J. 1 (1983); Michael K. Curtis, *Still Further Adventures of the Nine-Lived Cat: A Rebuttal to Raoul Berger's Reply on Application of the Bill of Rights to the States*, 62 N.C. L. REV. 517 (1984).

results. This approach may not eliminate all disagreement, but at least it sharpens the issues. Robert Hillman, for instance, used content analysis to challenge the emerging scholarly consensus that promissory estoppel was overtaking consideration as the basis for enforcing contracts.⁹⁰ This engendered several more case coding projects to refine the relevant data and deepen the debate about the meaning of documented trends.⁹¹ Even though coding and analyzing cases did not end the discussion, it moved the debate to a more observable evidentiary basis. Despite the lack of clear resolution, this series of articles on promissory estoppel serves as an encouraging example of the legal academy using classic social-scientific methods to critique and correct itself.

B. What Content Analysis Can and Cannot Tell Us

Because content analysis is better suited to answering certain research questions than others, the epistemological ambitions of the researcher will dictate whether and when to favor this method. In this Section, we explore the range of research questions that content analysis might examine. We consider the epistemological fit between this method and projects investigating (1) the bare outcomes of legal disputes, (2) the legal principles one can extrapolate from those outcomes, and (3) the facts and reasons that contribute to those outcomes and principles.⁹² Finally, we consider an ambitious set of studies, known as “jurimetrics,” that attempt to predict the impact of real-world facts on litigation. In the end, we conclude that this last use of content analysis overreaches the method’s epistemological aims. Fittingly, the legally-trained researchers who can best take advantage of the strengths of content analysis are those who ask questions similar to the questions that traditional legal scholars ask. The different methods produce distinctive insights, but they are well-suited to address similar issues.

1. Counting Case Outcomes

Researchers frequently analyze the bare outcomes rather than the content of opinions. They test non-legal explanations for judicial voting patterns or case outcomes, including political influences, judges’ personal characteristics, or institutional factors (such as court structure). Measuring case results differs fundamentally from measuring the law of a case. The narrower and more objective questions about who won and who lost require less legal judgment than distilling the legal principle a case embodies.

90. Hillman, *supra* note 27.

91. See, e.g., Juliet P. Kostritsky, *The Rise and Fall of Promissory Estoppel or is Promissory Estoppel Really as Unsuccessful as Scholars Say it is: A New Look at the Data*, 37 WAKE FOREST L. REV. 531 (2002).

92. See Hammer & Sage, *supra* note 27 and Bernard Trujillo, *Patterns in a Complex System: An Empirical Study of Valuation in Business Bankruptcy Cases*, 53 UCLA L. REV. 357 (2005) for similar, but somewhat different, categorizations of types of case coding studies.

The most prominent examples of this genre question the fundamental rule of law in judicial opinions. For over a decade, prominent political scientists, law professors, and judges have engaged in a heated debate over whether judges' decisions relate more to their personal or political views than to the neutral substance of the law and the facts of the case.⁹³ This debate has spawned a vast and often contentious literature, much of which is based on empirical studies using content analysis.⁹⁴ A calmer back-eddy of this type of research studies whether various institutional structures—such as elected versus appointed judges, their length of tenure, the use of specialized courts, crowded dockets, and the like—affect case outcomes or judicial reasoning.⁹⁵

Broader studies could include a wide range of external social and economic phenomena that might affect the content of judge-made law. Treating case outcomes as the dependent variable, the range of potential influences on judicial behavior that might be studied statistically is limited only by the bounds of a researcher's imagination.

A final category of analytic studies has been largely overlooked as a potential application for content analysis: those treating case law as an *independent* variable, meaning that they ask how case law influences other social and economic conditions. Law's effect on society is obviously a rich field of inquiry, but most such studies trace the effects only of statutory law.⁹⁶

93. For reviews and analysis, see Cross, *supra* note 27; George, *supra* note 72; Gillman, *supra* note 74; Sisk & Heise, *supra* note 2; Emerson H. Tiller & Frank B. Cross, *What is Legal Doctrine?* 100 NW. U. L. REV. 517 (2006).

94. See DAVID E. KLEIN, MAKING LAW IN THE UNITED STATES COURTS OF APPEALS (2002); Sara C. Benesh & Wendy L. Martinek, *State Supreme Court Decision Making in Confession Cases*, 23 JUSTICE SYSTEM JOURNAL 109 (2002); James J. Brudney, *A Famous Victory: Collective Bargaining Protections and the Statutory Aging Process*, 74 N.C. L. REV. 939 (1996); Cross & Tiller, *supra* note 27; Tracey E. George & Lee Epstein, *On the Nature of Supreme Court Decision Making*, 86 AM. POL. SCI. REV. 323 (1992); Charles A. Johnson, *Law, Politics, and Judicial Decision Making: Lower Federal Court Uses of Supreme Court Decisions*, 21 L. & SOC'Y REV. 325 (1987); Revesz, *supra* note 27; Daniel M. Schneider, *Empirical Research on Judicial Reasoning: Statutory Interpretation in Federal Tax Cases*, 31 N.M. L. REV. 325 (2001); Jeffrey A. Segal, *Predicting Supreme Court Cases Probabilistically: The Search and Seizure Cases, 1962-1981*, 78 AM. POL. SCI. REV. 891 (1984); Sisk et al., *supra* note 27; Sisk et al., *supra* note 47; Joseph L. Smith & Emerson H. Tiller, *The Strategy of Judging: Evidence from Administrative Law*, 31 J. LEGAL STUD. 61 (2002); Donald R. Songer & Susan Haire, *Integrating Alternative Approaches to the Study of Judicial Voting: Obscenity Cases in the U.S. Courts of Appeals*, 36 AM. J. POL. SCI. 963 (1992); Spaeth, *supra* note 24; Paul J. Wahlbeck, *The Development of a Legal Rule: The Federal Common Law of Public Nuisance*, 32 L. & SOC'Y REV. 613 (1998).

95. See KLEIN, *supra* note 94; Jason J. Czarnecki & William K. Ford, *The Phantom Philosophy? An Empirical Investigation of Legal Interpretation*, 65 MD. L. REV. 841 (2006); Gerard S. Gryski, Eleanor C. Main, & William J. Dixon, *Models of State High Court Decision Making in Sex Discrimination Cases*, 48 JOURNAL OF POLITICS 143 (2001); Grundfest & Pritchard, *supra* note 27; Melinda G. Hall & Paul Brace, *State Supreme Courts and Their Environments: Avenues to General Theories of Judicial Choice*, in SUPREME COURT DECISION-MAKING: NEW INSTITUTIONALIST APPROACHES (Cornell W. Clayton & Howard Gillman, eds.) (1999); Wagner & Petherbridge, *supra* note 27.

96. A leading and especially contentious example is whether gun control laws affect crime

Researchers have neglected the possible effects of judge-made law, including statutory interpretation. With its diverse “laboratory of states,” the United States offers boundless opportunities to learn from the “natural experiments” created by the inevitable differences in case law among jurisdictions and over time.⁹⁷ Doing so, however, requires accurate and replicable coding of the case law.

While these applications of content analysis to litigation outcomes are plentiful and powerful, they do not capture the method’s full potential to study the *legal* content of judicial opinions. For stronger connections between content analysis and traditional interpretive readings of judicial opinions, we set our sights on studies that move deeper into the facts and reasoning of opinions.

2. Evaluating Legal Doctrine

a. Judicial Reasoning

Legal scholars study cases not simply because they reflect or respond to the law, but mainly because they *are* the law. To know the law of a case classically requires subjective interpretation, sometimes of a deeply reflective nature. Legal readers ponder the meaning of a decision for future cases by asking how the outcome in the current case relates to its facts, procedural posture, and the court’s reasoning.⁹⁸ It is doubtful that content analysis standing alone will be able to replace this core mode of interpretive case law analysis.

We have seen, however, that content analysis can augment conventional analysis by identifying previously unnoticed patterns that warrant deeper study, or sometimes by correcting misimpressions based on ad hoc surveys of atypical cases. In doing so, content analysis reaches a thinner understanding of the law than that gained through more reflective and subjective interpretive methods. The coding of case content does not fully capture the strength of a particular judge’s rhetoric, the level of generality used to describe the issue, and other subtle clues about the precedential value of the opinion. To some extent, the method trades depth for breadth.

The method also trades the pretense of ontological certainty for a more provisional understanding of case law. Content analysis tolerates large elements of imperfection. As described in Parts III.A.2 and III.B.3, its methods require only acceptable levels of replicability and statistical association.⁹⁹ The epistemology behind much social science tells us that ultimate reality is subject

rates. See Ian Ayres & John Donohue III, *Shooting Down the “More Guns, Less Crime” Hypothesis*, 55 STAN. L. REV. 1193 (2003).

97. An important resource for such studies is Morriss, *supra* note 47.

98. See generally BENJAMIN N. CARDOZO, *THE NATURE OF THE JUDICIAL PROCESS* (1921).

99. See also KLEIN, *supra* note 94, at 55; Juliano & Schwab, *supra* note 76 at 557.

only to approximation. Conventional legal analysis, however, does not cede this epistemological ground. In a realist mode, the law is what judges do; positivistically, law is at least some of what judges say about what they are doing. Legal analysts may not agree on what judicial opinions mean, but law's interpretive method allows each reader to claim a fully accurate understanding of the law.¹⁰⁰

These two points of contrast between traditional and content analysis of judicial reasoning—the depth and the objectivity of understanding—intertwine. Social scientists often observe a trade-off between the usefulness of findings and a method's objectivity. More perfectly replicable methods tend to produce narrower, more superficial, and therefore less meaningful findings.¹⁰¹ Another way to express this contrast is between internal and external validity. Content analysis is internally valid if it accurately measures the particular components of the decision that the researcher wants to study. Systematic coding can increase internal validity by removing elements of researcher bias and improving thoroughness and accuracy. However, to the extent that content analysis ignores aspects of legal interpretation that are impossible to code objectively (for example, nuances related to infrequent or highly complex factual and procedural patterns), content analysis loses relevance, or external validity. When uniform coding cannot capture important details about idiosyncratic decisions, content analysis alone fails to measure what lawyers or scholars would consider a full and accurate statement of the law.

b. Normative Legal Analysis

Because of these methodological limitations, very few opinion-coding projects explicitly aim to evaluate the legal correctness of judicial opinions.¹⁰² Almost all projects we reviewed aim primarily for descriptive or analytical, rather than normative, goals. Certainly, content analysts may draw normative implications from what they observe,¹⁰³ but their coding of cases aims only to

100. See generally RONALD DWORKIN, *LAW'S EMPIRE* (1986).

101. See KRIPPENDORFF, *supra* note 3 at 212-13. Legal scholars have noted the same point, see, for example, Fuller, *supra* note 81 at 1626-28 (contrasting narrow prediction of case outcomes with deeper understanding of judges' human behavior and motivation).

102. There are, however, other types of empirical studies that evaluate the legal correctness of jury decisions. For instance, a series of studies evaluate whether the results in medical malpractice cases accord with professional medical judgment regarding causation and the standard of care. For reviews of these studies, see Philip G. Peters, Jr. *The Role of the Jury in Modern Malpractice Law*, 87 IOWA L. REV. 909 (2002); Neil Vidmar, *Are Juries Competent to Decide Liability in Tort Cases Involving Scientific/Medical Issues? Some Data from Medical Malpractice*, 43 EMORY L.J. 885 (1994). However, these studies often struggle with acceptable reliability of expert professional judgment. See, e.g., Karen L. Posner, et al., *Variation in Expert Opinion in Medical Malpractice Review*, 85 ANESTHESIOLOGY 1049 (1996) (finding only 37% agreement beyond chance for review of appropriateness of care in anesthesia malpractice claims).

103. See, e.g., Willy E. Rice, *Federal Courts and the Regulation of the Insurance Industry: An Empirical and Historical Analysis of Courts' Ineffectual Attempts to Harmonize Federal Antitrust, Arbitration, and Insolvency Statutes with the McCarran-Ferguson Act--1941-1993*, 43

document what judges do rather than to assess in a formal empirical manner how well they perform. Without an independent “gold standard” for what the law *should* be in any particular case or jurisdiction, who is to say judges are wrong, in an empirical sense? What they *is* the law.¹⁰⁴

Among the few instances of normative analysis, two address only the accuracy of citations. Karl Llewellyn in 1960 reviewed 588 random case citations in judicial opinions to determine whether they accurately distinguished between dicta and holdings (they did).¹⁰⁵ Similarly, two political scientists verified the accuracy and reliability of Shepard’s citation service using a random sample of 602 opinions.¹⁰⁶ Only one project squarely evaluates the legal correctness of judicial opinions. It examines whether lower courts correctly applied or interpreted the Supreme Court’s libel law precedent, *New York Times v. Sullivan*.¹⁰⁷ The author was a political scientist, not a lawyer, and he did not discuss the reliability or validity of his own evaluations of the courts’ decisions.¹⁰⁸

These limited forays should caution those who might consider using content analysis to evaluate the legal holdings of judicial decisions. Legally trained readers appreciate the difficulties in making normative legal evaluations objectively. Doing so would require showing that judging and legal interpretation can be done with scientific reproducibility—a task fraught with jurisprudential difficulties.¹⁰⁹ Considering these limitations, normative legal evaluation is convincing only when a project includes a traditional interpretive

CATH. U. L. REV. 399 (1993); Willy E. Rice, *Judicial Bias, The Insurance Industry and Consumer Protection: An Empirical Analysis of State Supreme Courts’ Bad-Faith, Breach-of-Contract, and Breach-of-Covenant-of-Good-Faith and Excess-Judgment Decisions, 1900-1991*, 41 CATH. U. L. REV. 325 (1992); Rice, *supra* note 58, at 598, 599, 616, 619, 674 (expressing “anger” and “outrage” at the “astonishing” results from content analysis of insurance cases, which found that “impermissible and prejudicial factors” influence courts’ decisions, which are “despairingly strained, contradictory, divisive, and unintelligible.”)

104. Moreover, thoroughly evaluating the correct result or reasoning would often require much more information than what is provided in a typical opinion. Ideally, researchers attempting this would need access to the full set of materials and information submitted to the court.

105. LLEWELLYN, *supra* note 22.

106. Spriggs & Hansford, *supra* note 13.

107. John Gruhl, *The Supreme Court’s Impact on the Law of Libel: Compliance by Lower Federal Courts*, 33 WEST. POL. Q. 502 (1980).

108. Two other political scientists each code for whether courts failed to follow or to consider applicable precedent, but neither make this the main focus of the study. See KLEIN, *supra* note 94, at 54; Johnson, *supra* note 94. The first author notes the large element of subjectivity in this coding but does not test its reliability. KLEIN, *supra* note 94. The second author reports that there were too few instances of unambiguous noncompliance with precedent to include them in the analysis. Johnson, *supra* note 94 at 327.

109. It is interesting to contrast the field of clinical medicine, where researchers have developed replicable, scientific methods for evaluating the “appropriateness” of treatment decisions, using panels of expert physicians. See *infra* note 224. This is equivalent, however, only to evaluating the quality of professional legal advice, not the correctness of judicial decisions. Devising a convincing means to set an expert “gold standard” for judging is an entirely different matter that is conceptually much more challenging.

component.

3. *Exploring the Landscape of Case Law*

We face a Goldilocks dilemma—one where the best use of content analysis is not to aim for too much, or too little, but just enough insight. Content analysis does not give us the deepest understanding of judicial decisions and so limits our ability to discern or critique governing legal principles in ways that lawyers and judges find most relevant. However, studying opinions simply as vessels for bare outcomes or case holdings, while insightful, is not fully satisfying because such studies do not take full advantage of the rich reservoir of information within judicial opinions. It would be a waste to study only the skin of cases and to throw away their fruit. Judicial opinions are detailed repositories that show what kinds of disputes come before courts, how the parties frame their disputes, and how judges reason to their conclusions. It is the factual and analytical richness of judicial opinions that establish their substantive legal importance. Empirical legal researchers correctly recognize that it “is almost impossible to study law in a meaningful way without some attention to the [content of] opinions that contain these justifications.”¹¹⁰

Accordingly, scholars generally use content analysis in descriptive or explanatory studies. Most projects we reviewed attempt to take full advantage of the substance of judicial opinions, while avoiding normative legal analysis, by using factual, rhetorical, and legal details to describe or explore a body of case law, a legal doctrine, or an aspect of opinion-writing. The approach is loosely structured, calling on the researcher simply to observe and document what can be found, as a naturalist might explore a new continent or even a familiar patch of woods by turning over stones to see what crawls out. This general approach contrasts with more focused analytic projects that use formal, statistical hypothesis testing to generate definitive conclusions about cause-effect relationships that have theoretical significance. Descriptive or exploratory studies are more akin to mapping than to testing.

There are two broad subdivisions of descriptive/exploratory studies: (1) those that examine the background of legal doctrines, case subject matter, or case outcomes, versus (2) those that focus on particular techniques of opinion-writing, such as syntax, semantics, citations, or reasoning style. The first type is the most common. These researchers code cases to document trends in the case law and the factors that appear important to case outcomes. A dozen explore a variety of civil rights topics,¹¹¹ another dozen explore various business law

110. Barry Friedman, *Taking Law Seriously*, 4 PERSPECTIVES ON POLITICS 261, 266 (2006). *Accord* Epstein et al., *supra* note 78 at 320-23 (explaining why coding projects should focus on more than mere case outcomes).

111. See Ruth Colker, *Winning and Losing under the Americans with Disabilities Act*, 62 OH. ST. L. J. 239 (2001); Eisenberg & Johnson, *supra* note 76; Juliano & Schwab, *supra* note 76;

topics,¹¹² and a half dozen others address issues in health care law.¹¹³ Many focus on who wins and loses and why, but others simply want to understand what is happening in the courts.¹¹⁴ Some ask about the impact of a new precedent, statute, or legal doctrine.¹¹⁵ Most prominent among these are a passel of articles exploring the impact of the Supreme Court's *Chevron*

Victor E. Kappeler, Stephen F. Kappeler, & Rolando V. del Carmen, *A Content Analysis of Police Civil Liability Cases: Decisions of the Federal District Courts, 1978-1990*, 21 JOURNAL OF CRIMINAL JUSTICE 325 (1993); Willy E. Rice, *Judicial and Administrative Enforcement of Title VI, Title IX, and Section 504: A Pre- and Post-Grove City Analysis*, 5 REV. LITIG. 219 (1986); Willy E. Rice, *Judicial Enforcement of Fair Housing Laws: An Analysis of Some Unexamined Problems That the Fair Housing Act of 1983 Would Eliminate*, 27 HOW. L. J. 227 (1984); Darrell L. Ross, *Emerging Trends in Correctional Civil Liability Cases: A Content Analysis of Federal Court Decisions of Title 42 United States Code Section 1983: 1970-1994*, 25 J. OF CRIMINAL JUSTICE 501 (1997); Kathryn E. Scarborough & Craig Hemmens, *Section 1983 Suits Against Law Enforcement in the Circuit Courts of Appeal*, 21 T. JEFFERSON L. REV. 1 (1999); Schultz & Petterson, *supra* note 38; Michael Schuster, *Analyzing Age Discrimination Act Cases: Development of a Methodology*, 4 LAW & POL'Y Q. 339 (1982); Adam Winkler, *Fatal in Theory and Strict in Fact: An Empirical Analysis of Strict Scrutiny in the Federal Courts*, 59 VAND. L. REV. 793 (2006).

112. See Wayne N. Bradley, Comment, *An Empirical Study of Defective Incorporation*, 39 EMORY L.J. 523 (1990); Theresa Gabaldon, *A Sense of Security: An Empirical Study*, 25 J. CORPORATION L. 307 (2000); Christopher C. Klein, *Predation in the Courts: Legal Versus Economic Analyses in Sham Litigation Cases*, 10 INT'L REV. L. & ECON. 29 (1990); Rice, *Federal Courts and the Regulation of the Insurance Industry*, *supra* note 103; Willy E. Rice, *Insurance Contracts and Judicial Decisions over Whether Insurers Must Defend Insureds that Violate Constitutional and Civil Rights: An Historical and Empirical Review of Federal and State Court Declaratory Judgments 1900-2000*, 35 TORT & INS. L. J. 995 (2000); Willy E. Rice, *Insurance Contracts and Judicial Discord over Whether Liability Insurers Must Defend Insureds' Allegedly Intentional and Immoral Conduct: A Historical and Empirical Review of Federal and State Courts' Declaratory Judgments 1900-1997*, 47 AM. U.L. REV. 1131 (1998); Willy E. Rice, *Insurance Decisions: A Survey and an Empirical Analysis*, 35 TEX. TECH. L. REV. 947 (2004); Rice, *Judicial Bias*, *supra* note 103; Rice, *Race, Gender "Redlining,"* *supra* note 58; Alton L. Absher, III & Steven N. Baker, *Empirical Study, Unconscionability Never Really Works, Does It? An Empirical Study of How Courts Have Treated Unconscionability Under UCC Section 2-302*, WAKE FOREST L. REV. Working Paper No. 4 (2005), available at <http://www.law.wfu.edu/prebuilt/Absher-Baker-ES.pdf>.

113. See Hall et al., *supra* note 28; Hammer & Sage, *supra* note 27; Jacobsen et al., *supra* note 76; Edward Alan Miller, *Federal Oversight, State Policy Making, and the Courts: An Empirical Analysis of Nursing Facility Litigation Under the Boren Amendment*, 3 J. EMPIRICAL LEG. STUD. 145 (2006); Sandor, *supra* note 19; Stetler, *supra* note 19.

114. For instance, see Gabaldon, *supra* note 112, for how courts define securities, and Kagan et al., *supra* note 24, for what kinds of cases have state supreme courts decided over the course of a century.

115. See Bradley, *supra* note 112; Lloyd Dixon & Brian Gill, *Changes in the Standards for Admitting Expert Evidence in Federal Civil Cases Since the Daubert Decision*, 8 PSYCHO. PUB. POL'Y L. 251 (2002); Eisenberg & Johnson, *supra* note 76; Fradella et al., *supra* note 70; Franklin, *supra* note 24; Marsha Garrison, *How Do Judges Decide Divorce Cases? An Empirical Analysis of Discretionary Decision Making*, 74 N.C. L. REV. 401 (1996); Susan Gluck Mezey & Susan M. Olson, *Fee Shifting and Public Policy: The Equal Access to Justice Act*, 77 JUDICATURE 13 (1993); Jennifer L. Groscup et al., *The Effects of Daubert on the Admissibility of Expert Testimony in State and Federal Criminal Cases*, 8 PSYCH. PUB. POL'Y & L. 339 (2002); Gruhl, *supra* note 107; Solimine, *supra* note 87.

decision regarding judicial deference to administrative agency decisions.¹¹⁶

An especially interesting subgenre uses content analysis to find some order and logic in a body of case law that, by conventional analysis, appears chaotic or haphazard. As Fred McChesney notes, “[t]he academic history of American law generally is replete with instances in which scholars have proclaimed traditional common law modes of distilling ‘the law’ from cases unworkable.”¹¹⁷ These conventional legal analysts, throwing up their hands, conclude that the law on the topic is hopelessly confused and inconsistent, or less pejoratively, dependent on individual facts. Leading examples of these unwieldy topics include obscenity law, unreasonable search and seizure, unconscionability of contracts, and piercing the corporate veil. Content analysts have risen to the challenge of finding some hidden order in each of these bodies of otherwise unfathomable case law.¹¹⁸

Critics of these descriptive or exploratory studies contend that they sometimes draw conclusions about features of the legal landscape that one cannot fully observe from judicial opinions. As discussed in Part III-A, win/loss records from published opinions do not necessarily tell us about legal disputes that were never filed in court, those that the parties settled, or those that judges resolved without written or published opinions. Nevertheless, even if judicial opinions offer a skewed view of what occurs elsewhere in the system, they are a highly valuable source for systematic study, revealing the portion of the legal world that, in many ways, is most important. As Lawrence Friedman and colleagues explain, “[i]n the theory of the common law, these opinions *are* the law; they stand in the center of the legal system. Their power is enhanced by the common law doctrine that links them in a chain of influence and causation—the doctrine of precedent.”¹¹⁹ Therefore, as Bernard Trujillo puts it, published opinions are especially useful for studying the spread of ideas within the legal system:

116. See Cross & Tiller, *supra* note 27; Orin S. Kerr, *Shedding Light on Chevron: An Empirical Study of the Chevron Doctrine in the U.S. Courts of Appeals*, 15 YALE J. ON REG. 1 (1998); Merrill, *supra* note 27; Revesz, *supra* note 27; Mark J. Richards, Joseph L. Smith & Herbert M. Kritzer, *Does Chevron Matter?* 28 LAW & POL’Y 444 (2006); Christopher H. Schroeder & Robert L. Glicksman, *Chevron, State Farm, and EPA in the Courts of Appeals During the 1990s*, 31 ENVTL. L. REP. 10371 (2001); Schuck & Elliot, *supra* note 25.

117. McChesney, *supra* note 76, at 533.

118. See *id.* (defective incorporation); FRED KORT, CONTENT ANALYSIS OF JUDICIAL OPINIONS AND RULES OF LAW, in JUDICIAL DECISION-MAKING (Glendon Schubert ed., 1963) (involuntary confession); Beebe, *supra* note 27 (trademark infringement); Haar et al., *supra* note 7 (zoning variances); Timothy M. Hagle, *But Do They Have to See It to Know It? The Supreme Court’s Obscenity and Pornography Decisions*, 44 W. POL. Q. 1039 (1991) (obscenity); Kort, *supra* note 13 (right to counsel); Kevin T. McGuire, *Obscenity, Libertarian Values, and Decision Making in the Supreme Court*, 18 AMER. POLITICS Q. 47 (1990); Ostas, *supra* note 58 (unconscionability); Segal, *supra* note 94 (search and seizure); Robert B. Thompson, *Piercing the Corporate Veil: An Empirical Study*, 76 CORNELL L. REV. 1036 (1991) (piercing the corporate veil); Trujillo, *supra* note 92 (valuation of assets in bankruptcy).

119. See Friedman, et al., *supra* note 72, at 773. *Accord* Friedman, *supra* note 110, at 266.

[P]ublished opinions are an important “communications device” that travel among the elements of the system, like proteins in a cell. Judges intend their published opinions not only as a communication to the parties in the particular case that gave rise to the opinion, but also as a communication to other judges, other lawyers, other litigants, and other actual and potential participants in the legal system.¹²⁰

Accordingly, content analysis is perfectly suited for examining aspects of judicial *method*. Questions researchers have pursued include the types of authorities judges cite in their opinions;¹²¹ the argumentative, interpretive, or expressive techniques judges use in different circumstances;¹²² and the various meanings of important social conceptions (such as family or equality) that their written opinions tend to reflect.¹²³ This category includes some of the most inventive uses of case content analysis, such as Laura Little’s study of linguistic devices used to obscure meaning in Supreme Court opinions.¹²⁴

The largest grouping in this general category is the half dozen authors who analyze courts’ reliance on social policy or social science, in contrast with more technical or formalistic doctrinal reasoning.¹²⁵ Naturally, caution is warranted in concluding that a mention of some source in an opinion indicates the actual importance that judges place on this type of evidence and argument. Still, with appropriate caveats on the claims being made, systematic study of how judges reason in their written decisions is a compelling application of case content analysis because it best fits the method with the type of question that researchers are asking.

120. Trujillo, *supra* note 92, at 364-65.

121. See KLEIN, *supra* note 94; LLEWELLYN, *supra* note 22; James R. Acker, *A Different Agenda: The Supreme Court, Empirical Research Evidence, and Capital Punishment Decisions, 1986-1989*, 27 LAW & SOC’Y REV. 65 (1993); Bernstein, *supra* note 21; Brudney & Ditslear, *supra* note 76; Davis, *supra* note 25; Patricia J. Falk, *The Prevalence of Social Science in Gay Rights Cases: The Synergistic Influences of Historical Context, Justificatory Citation, and Dissemination Efforts*, 41 WAYNE L. REV. 1 (1994); Henry F. Fradella, *A Content Analysis of Federal Judicial Views of the Social Science “Researcher’s Black Arts,”* 35 RUTGERS L. J. 103 (2003); Susan Hedman, *Friends of the Earth and Friends of the Court: Assessing the Impact of Interest Group Amici Curiae in Environmental Cases Decided by the Supreme Court*, 10 VA. ENVTL. L.J. 187 (1991).

122. See Czamezki & Ford, *supra* note 95; Daynard, *supra* note 21; Epstein et al., *supra* note 78; Henderson & Eisenberg, *supra* note 51; Charles A. Johnson, *Content-Analytic Techniques and Judicial Research*, 15 AM. POLITICS Q. 169 (1987); Little, *supra* note 50; Glenn A. Phelps & John B. Gates, *The Myth of Jurisprudence: Interpretive Theory in the Constitutional Opinions of Justices Rehnquist and Brennan*, 31 SANTA CLARA L. REV. 567 (1991); Rebell & Block, *supra* note 11; Schacter, *supra* note 27; Schneider, *supra* note 94; Sisk et al., *supra* note 27; Smith & Tiller, *supra* note 94; Wagner & Petherbridge, *supra* note 27.

123. See Vicki Lens, *Supreme Court Narratives on Equality and Gender Discrimination in Employment: 1971-2002*, 10 CARDOZO WOMEN’S L.J. 501 (2004); Richman, *supra* note 58.

124. Little, *supra* note 50.

125. See Brudney & Ditslear, *supra* note 121; Davis, *supra* note 25; Daynard, *supra* note 21; Falk, *supra* note 121; Fradella, *supra* note 121; Phelps & Gates, *supra* note 122; Rebell & Block, *supra* note 11.

4. *The Special Case of Jurimetrics*

The most ambitious use of content analysis is to study the factors that determine the outcomes of cases. Researchers commonly explore the possible connections between case results and various explanatory factors.¹²⁶ Some researchers, though, pursue a more structured and ambitious goal. They use sophisticated statistical methods to formally model or predict judges' behavior, the content of opinions, or relationships to external factors. This general approach was at one time called "jurimetrics."¹²⁷ This kind of study might, for instance, attempt to predict the likely result in a case when the parties present the judge with a particular combination of legally relevant factors. Often, the stated purpose is to help practicing lawyers make better-informed decisions about handling particular cases.¹²⁸ Other times, the purpose is more scholarly—to test various claims based on legal theory.¹²⁹

Jurimetrics draws the most pointed contrast between content analysis and

126. For instance, many projects study the correlation with factors that should, in theory, be irrelevant to the court's decision, such as the judge's political affiliation or the resources available to the litigants. See *supra* text accompanying notes 93-95.

127. See, e.g., Mendelson, *supra* note 75; Glendon Schubert, *The Importance of Computer Technology to Political Science Research in Judicial Behavior*, 8 JURIMETRICS J. 56 (1968). "Jurimetrics" also referred to scientific study of legal issues more generally, with an emphasis on the novel uses of computers. See generally Nicholas Johnson, *Jurimetrics and the Association of American Law Schools*, 14 J. LEG. ED. 385 (1962) (surveying multiple uses of the term); Lee Loevinger, *Jurimetrics: Science and Prediction in the Field of Law*, 46 MINN. L. REV. 255 (1961) (reviewing development of the concept); Symposium, *Jurimetrics*, 28 LAW & CONTEMP. PROBS. 1 (1963) (collecting various applications of the concept). For a recent application, see Kevin D. Ashley & Stefanie Brüninghaus, *Computer Models for Legal Prediction*, 46 JURIMETRICS J. 309 (2006).

128. Charles Haar and colleagues' early 1970s work on state zoning cases is still the most refined example of this. Haar et al., *supra* note 7. They use sophisticated statistical analyses to generate a table that allows lawyers to calculate precise probability predictions of winning a case based on the combination of discrete factors involved, considering each factor either in isolation, or cumulatively. *Id.* at 657. For less elaborate efforts to advise practicing lawyers, see, for example, Absher & Baker, *supra* note 112; Allison & Lemley, *supra* note 87.

129. See, SARA C. BENESH, THE U.S. COURT OF APPEALS AND THE LAW OF CONFESSIONS: PERSPECTIVES ON THE HIERARCHY OF JUSTICE (2002); Beebe, *supra* note 27; Benesh & Martinek, *supra* note 94; Antony W. Dnes & Jonathan S. Seaton, *An Exploration of the Tort-Criminal Boundary Using Manslaughter and Negligence Cases*, 17 INT'L REV. L. & ECON. 537 (1997); Eisenberg & Johnson, *supra* note 76; Grunbaum & Newhouse, *supra* note 20; Gryski et al., *supra* note 95; Henderson & Eisenberg, *supra* note 51; Johnston, *supra* note 27; Kerr, *supra* note 116; Herbert M. Kritzer & Mark J. Richards, *The Influence of Law in the Supreme Court's Search-and-Seizure Jurisprudence*, 33 AMER. POLITICS RESEARCH 33 (2005); McChesney, *supra* note 58; Mark J. Richards & Herbert M. Kritzer, *Jurisprudential Regimes in Supreme Court Decision Making*, 96 AM. POL. SCI. REV. 305 (2002); Herbert M. Kritzer & Mark J. Richards, *Jurisprudential Regimes and Supreme Court Decision Making: The Lemon Regime and Establishment Clause Cases*, 37 L. & SOC'Y REV. 827 (2003); David Sherwyn, Michael Heise, & Zev J. Eigen, *Don't Train Your Employees and Cancel Your "1-800" Harassment Hotline: An Empirical Examination and Correction of the Flaws in the Affirmative Defense to Sexual Harassment Charges*, 69 FORDHAM L. REV. 1265 (2001); Songer & Haire, *supra* note 94; John A. Swain & Edwin E. Aguilar, *Piercing the Veil to Assert Personal Jurisdiction Over Corporate Affiliates: An Empirical Study of the Cannon Doctrine*, 84 B.U. L. REV. 445 (2004).

more traditional interpretive methods. Scholars in the jurimetrics tradition seek to test precise hypotheses or construct formal models of how complex legal or factual elements affect case outcomes.¹³⁰ Charles Haar, an early and overlooked visionary of this method, underscores the potential advantages over traditional legal analysis:

Legal precedents hidden in the corners of recorded cases are scrutinized as closely as were the entrails of pigeons by the high priests of Apollo. Lawyers expend time and energy probing case law, searching for patterns, discerning trends, then struggling to impress the court with the similarities or differences between their clients' cases and the cherished legal precedents. . . . Yet this process—in crucial aspects—is no less mysterious than that of the Greek seers and reflects as much intuition resulting from personal immersion in the cases, as it does a rational process of analysis and dissection. This is an unhappy state of affairs. Need the profession continue to struggle in splendid isolation to decipher the future? Just as other areas of the social sciences have turned to [statistical] analysis for assistance in sorting and analyzing complex sets of data, so could law.¹³¹

This aspiration encounters the major analytical difficulty, however, that the facts gleaned from judicial opinions are likely to be both incomplete and biased.

a. The Circularity of Facts in Judicial Opinions

Analyzing the cause-and-effect relationship between legally relevant factors and the content of judicial opinions raises a serious circularity problem: the facts and reasons found in an opinion might or might not accurately describe the real world facts or the true nature of the judge's decision-making process. There is no reason to expect that appellate opinions should provide complete, objective, and result-neutral statements of all the facts in each case. Instead, there is every reason to think just the opposite. Therefore, content analysts must acknowledge that a "judicial opinion is the judge's story justifying the judgment. The cynical legal realist might say that the facts the judge chooses to relate are inherently selective and a biased subset of the actual facts of the case."¹³²

This limitation entails two distinct problems: factual incompleteness, and factual distortion.¹³³ Incompleteness results because judges' factual

130. For example, this was Fred Kort's aim in the first conscious application of content analysis to appellate cases. As described *supra*, in text at notes 12-13, Kort showed that the outcome of Supreme Court right-to-counsel cases could be predicted reliably from various combinations of factors stated in the opinions.

131. Haar et al., *supra* note 7, at 655.

132. Juliano & Schwab, *supra* note 76, at 558-559. See also Krawiec & Zeiler, *supra* note 27, at 1832 (discussing this concern).

133. These also can be thought of as Type II and Type I errors, respectively, or, in testing

presentations are meant only to explain as much of the case as is necessary to justify the outcome.¹³⁴ Judicial parsimony can severely distort analysts' measurement of important facts. Juliano and Schwab provide an apt example from their study of sexual harassment cases. In coding for plaintiffs' race, they identified whites in only about one quarter of cases where race could be determined.¹³⁵ Most likely, this results from courts' reluctance to mention race when not legally relevant in a particular case, and race is less often legally relevant for whites. Social scientists who want to study the role of race across a body of case law, however, need a reasonably accurate indication in every case, not just in those cases in which the judge considered race legally relevant.

The second problem is the possibility that judges distort the facts they report to justify the legal results they reach. While this is a highly contentious charge, distortion does not have to amount to an outright misrepresentation of facts. Instead, distortion results simply when judges emphasize opposing facts less than supporting facts. Even if this much is true—that the facts used to statistically predict outcomes are those that the courts select and filter to justify their results—then this analytical method faces a very serious challenge indeed. It is circular or tautological to predict judicial outcomes from facts that *reflect* rather than *generate* the result.¹³⁶ Conclusions drawn from statistical analysis of facts reported in this manner will not aid a lawyer or scholar who wants to predict the outcome of future cases or to understand what truly gives rise to a body of precedent.

Defenders of jurimetrics observe that the incomplete and distorted view of the facts from judicial opinions is "precisely what every lawyer [uses] when

jargon, false negatives and false positives: falsely assuming facts are absent from the case when they are not mentioned and wrongly assuming that the facts described are true.

134. Reed C. Lawlor, *Fact Content Analysis of Judicial Opinions*, 8 JURIMETRICS J. 107, 109-10 (1968).

135. Juliano & Schwab, *supra* note 76, at 558.

136. See Mendelson, *supra* note 74, at 601-02; William M. Sage, *Judicial Opinions Involving Health Insurance Coverage: Trompe L'Oeil or Window on the World?*, 31 IND. L. REV. 49, 66-67 (1998) ("[S]ome stated [judicial] rationales are fabrications intended to clothe otherwise naked truth, drawing empirical conclusions from them may be hazardous."). Lon Fuller made a similar critique in commenting on Kort's analysis (described in text at notes 12-13 *supra*):

There are serious difficulties in Kort's proposed method It presupposes a clear distinction between findings of fact and conclusions of law that cannot be drawn. Kort suggests as one variable that may be present in criminal cases, "the alleged fact that the defendant had not been advised of his right to remain silent." Now, "being advised" must mean "being adequately advised." Whether a man is deemed adequately advised involves a conclusion of law. Kort recognizes that his method is "not designed to predict doctrinal changes and the adoption of new rules of law." But a "new rule of law" may quietly emerge by the simple process of tightening up the definition of what will justify a factual finding, let us say, that the defendant was sufficiently warned of his rights All of this can be summed up in the observation that the "findings of fact" Kort's method would feed into the computers will inevitably carry with them a heavy, but essentially unmeasurable, contamination of law.

Fuller, *supra* note 81, at 1615.

reading a case . . . ; the application of mathematical techniques cannot make the method less valid.”¹³⁷ But statistics cannot make this method of case law analysis *more* valid, or valid *at all*. As the quip goes in the world of computers: garbage in, garbage out. This serious flaw hidden in all conventional doctrinal analysis—one that critical legal theorists have taken great pains to uncover—is equally troubling for statistical as for conventional legal analysis.

b. Answering the Skeptics

We offer four responses to judicial skeptics, although these responses do not entirely remove the cloud of circularity. First, social science data does not need to be perfect; reasonable approximations are good enough for both government work and empirical work. Social scientists often survey the public about a variety of self-reported attitudes and behaviors—for instance, candidate preference and likelihood of voting—with the assumption that people answer these questions accurately and independently. Too often, this assumption is not true: people sometimes fail to say what they really think, say what they imagine the researcher wants to hear, or try to maintain logical consistency across questions even if this distorts the truth. Still, imperfect data must suffice because observing actual behaviors and gauging true attitudes would be impossible or cost-prohibitive. Similarly, even though judge-reported facts may not “purport to be the real facts,” they are “near enough so that the savings in labor justifies the approximation.”¹³⁸ This assumption is not heroic. The lawyers and law professors who stake their life’s work on believing (for the most part) judges’ renditions of facts are hardly naive idealists.

Second, researchers can examine the fidelity of reported facts, looking for indications of distortion or incompleteness, to determine if the facts are close enough to reality for use in statistical analysis.¹³⁹ One interesting technique,

137. Tyree, *supra* note 80, at 2. *Accord* Theodore Eisenberg & James A. Henderson, *Products Liability Cases on Appeal: An Empirical Study*, 16 JUSTICE SYS. J. 117, 188-190 (1993); Hillman, *supra* note 27, at 583-84. Charles Haar adds that coding is more objective because, unlike traditional legal analysis, it regards all facts as having equal weight regardless of whether the court attaches any importance to them in even whether it disclaims any or the opposite importance. Haar et al., *supra* note 7, at 743-45. Haar acknowledges, however, that this does not deal with the problem that courts may simply omit facts the researcher considers relevant, or may distort facts to support the court’s conclusions. For this, he falls back on the observation that the statistical researcher is no worse off than conventional legal analysts, but the statistician has at her disposal a much stronger set of analytical tools that account for a “multiplicity of variables across a far greater number of cases.” *Id.* at 744.

138. Tyree, *supra* note 80 at 32. *Accord* Lawlor, *supra* note 134 at 111, 117 (explaining the value of reported facts, even if we cannot determine “what the facts were in nature”, because a “complete examination of the record would be horrendous”).

139. Hillman, for instance, in his prominent study of promissory estoppel, closely read a sample of cases looking for any indications that courts stated there was reliance when there really was not. See Hillman, *supra* note 27. He reports there was little evidence of misrepresentation of the facts, but he does not explain how he made this determination or why he felt he should be able to detect misrepresentation. *Id.* Krawiec and Zeiler had students code cases both for whether the

used in four studies,¹⁴⁰ compares facts reported in an appellate opinion with those reported in either the trial court's opinion or a dissenting opinion. These researchers found few instances of disagreement or distortion.¹⁴¹ Such reliance on the adversarial process—here, competing accounts of the facts that appear in dissenting opinions—extends the lawyer's classic process-based technique for truth-finding at trial into the research arena.

Third, the "bias" created when courts justify their decisions may be precisely what a researcher wishes to study.¹⁴² After all, the facts and reasons the judge selects are the substance of an opinion that creates law and binding precedent, so they merit careful study for this very reason. However, this justification calls for precision in setting the goals of study. Instead of predicting outcomes, content analysis is better suited to studying judicial reasoning itself, retrospectively. Scholars can use the method to learn more, for instance, about how results are justified. This type of study may be less relevant to practicing lawyers trying to gauge their cases' likely outcomes, but it is perhaps more relevant to legal scholars seeking a measurable understanding of substantive law or the legal process.

The final justification for basing statistical analysis on imperfectly reported facts is that, despite its flaws, this source of data is simply too valuable to disregard entirely. Social scientists recognize that judicial opinions "are a tremendous resource,"¹⁴³ "probably the best single source of systematic historical data on the law."¹⁴⁴ Three-quarters of a century ago, Columbia Law Professor Herman Oliphant in his inaugural address as President of the

judge indicated the factor was present, and whether the student thought it was present, in order to "minimize the impact of conscious or unconscious judicial mischaracterizations of fact." Krawiec & Zeiler, *supra* note 27 at 1834. Only "rarely" did they find disagreement between the judge and the student coder. *Id.* at 1835.

140. See Joseph A. Ignagni, *U.S. Supreme Court Decision Making and the Free Exercise Clause*, 55 REV. POLITICS 511 (1993); Lawlor, *supra* note 134 at 115; Posner, *supra* note 11; Richards & Kritzer *Am. Pol. Sci. Rev.*, *supra* note 129, at 317.

141. Lawlor, *supra* note 134 at 115, is especially thorough. Using student coders, he tested for thirty possible facts in two Supreme Court cases that each had three opinions. Out of these sixty observations that provided one hundred eighty possibilities for discrepancies to appear, he found only two.

142. See Friedman, *supra* note 110 at 268-29 (June 2006) ("Skepticism over whether reported facts are the real facts is justified. But, the mistake is to believe that the facts therefore aren't meaningful. Instead, they reflect in part what the law is. So, reported facts can be analyzed as legal factors."); Trujillo, *supra* note 92 at 366 ("Bankruptcy valuation doctrine is useful evidence for studying system dynamics precisely because it is not an unvarnished report of an objective event—what was the found value of the asset?—but rather a subjective account of an objective event—what did the judge say about finding the value of the asset?"); Wagner & Petherbridge, *supra* note 27 (explaining that judges' theories of patent claim construction are best determined by how they reason their opinions because that is how they convey their sense of the law to the legal community). Offering similar justifications, see also Hillman, *supra* note 27; Ignagni, *supra* note 140.

143. Friedman et al., *supra* note 72 at 773.

144. Bliss Cartwright, *Conclusion: Disputes and Reported Cases*, 9 L. & Soc'y REV. 369, 383 (1975).

American Association of Law Schools exhorted the legal academy to make greater use of the “gold mine” of material in cases for legal research, “equaling [the amount of information] to which geology, for instance, has had access, and the individual records are not more fractional or otherwise imperfect.”¹⁴⁵ A half century ago, Karl Llewellyn recognized that the case reporters present an “amazing gathered treasure . . . waiting for explorations.”¹⁴⁶ Modern day legal researchers tend to agree, judging from how many have devoted considerable effort to using this data source.

In sum, the circularity problem represents the most serious of the ground-level objections to content analysis. Researchers should not assume that judicial opinions contain a complete and objective account of all the relevant research factors. This limits the usefulness of content analysis for lawyers who want to predict the likely outcome of litigation or an appeal based on real-world or trial-record views of the facts. But, this limitation is not fundamentally different from similar limitations in other fields of social science study. Few social scientists use content analysis to draw definitive cause-and-effect conclusions about complex events. Instead, they more often identify apparent associations of interest meriting further study.

Moreover, the same concerns about factual fidelity apply equally to traditional interpretive readings of judicial opinions. Both methods face difficulties if researchers assume that the facts mentioned in the opinions correspond to the facts on the ground. Both the interpretive and content analysis methods should use case facts mainly to identify which factors judges treat as relevant in various bodies of law. We can study what judges say and do, internal to the case law, even if we cannot measure externally what set of raw facts produced those legal results.

C. Summary

Content analysis is a valuable research tool for more rigorous study of the empirical claims raised in conventional legal analysis. It is a more systematic and objective way to document what courts do and what they say. The insights gained from uniform content analysis of large numbers of opinions supplement the deeper understanding of individual opinions that comes from traditional interpretive techniques. The content of judicial opinions can be important in the study of the broader social, economic, and political systems that interact with judicial precedent, but cases are also well worth scientific study in their own right.

Social science methods complement conventional legal methods in two general ways. Content analysis can verify or refute descriptions of case law that

145. Oliphant, *supra* note 1 at 161. *Accord* Trujillo, *supra* note 92 at 363 (stressing the value of the “large amounts of easily detectable data” produced by the bankruptcy courts).

146. LLEWELLYN, *supra* note 22 at 514.

are based on more anecdotal or subjective study. More importantly, content analysis can identify surface patterns (which are sometimes hidden from the naked eye of casual readers), to be explored more deeply through interpretive, theoretical, or normative legal analysis. Each method allows us to see something different about the behavior of judges and the content of case law.

The major limitation of content analysis—a limit that applies equally to traditional interpretive methods—is that one cannot treat as accurate and complete the facts and reasons given in opinions. Therefore, researchers should be cautious about the meanings they attach to observations made through content analysis. This method is not well suited for predicting with assurance which cases will be won or lost.¹⁴⁷ But it can tell us how cases should be developed and argued. It can also describe more accurately the landscape of decided cases, and it is the most precise way for documenting what appellate judges decide and how they explain their decisions.

Despite some overlap, content analysis does not occupy the same epistemological ground as conventional legal analysis. The interpretive method provides a nuanced understanding that is better suited for evaluation of legal principles. Content analysis, on the other hand, holds more promise in the study of the connections between judicial opinions and other parts of the social, political, or economic landscape. The strongest application, however, is when the subject of study is the behavior of judges in writing opinions. There, content analysis combines the analytical skills of the lawyer with the power of science that comes from articulated and replicable methods of reading and counting cases.

III METHODOLOGY GUIDELINES

So far, we have considered whether and when to use content analysis. In this part, we consider how best to construct a content analysis. Gregory Sisk and Michael Heise capture the spirit in which we offer methodological guidance for coding:

No study could survive a scrutiny that demands absolute perfection, in methodology or in expression of the results. For these reasons, it is a mistake, in our judgment, to speak too easily and too absolutely of *rules* that should govern legal empirical research. . . . [W]e suggest that the methodological discussion is better advanced by thinking of most of these suggested “rules” as *guidelines* or *standards*, that is, aspirations toward which one always strives but some of which cannot

147. But, then again, neither are conventional legal methods. In fact, statistical modeling appears to predict Supreme Court outcomes somewhat better than the subjective views of legal scholars. See Theodore W. Ruger, et al., *The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking*, 104 COLUM. L. REV. 1150 (2004).

be executed in a particular study. . . .

Weaknesses in measurement, even when unavoidable, necessarily weaken our confidence in the results. The answer is neither to forgo the study while waiting in vain for the day when perfection can be attained nor to conduct the study and announce the results as final and infallible. . . . We should work together as a community to develop *guidelines* that are more or less flexible depending upon the subject of study and the real-world considerations of available information, resources and time, etc.¹⁴⁸

Our guidelines are based on a review of the universe of case-coding projects (summarized in Table 2), recommendations from leading content analysis methodologists,¹⁴⁹ and our own experience in coding and counting cases.¹⁵⁰ We revisit the same three stages of content analysis discussed in Part II.A — selecting, coding, and analyzing cases — but we now focus on *how* to rather than *whether* to undertake each task.

Table 2: Characteristics of Case Coding Projects

| Percent of Projects With: | |
|----------------------------------|----|
| Replicable Case Selection | 87 |
| Universal Sample | 85 |
| Student Coding | 22 |
| Specific Coding Instructions | 60 |
| Reliability Testing | 14 |
| Only Descriptive Statistics | 34 |

n=134

A. Selecting Cases

1. Sampling Frame and Biases

An empirical researcher must first decide which cases to select and sample. There are two components to consider: sampling frame, and selection method. The *sampling frame* is the theoretical universe of all relevant cases, and the *selection method* determines which cases will actually be sampled and studied.

The selection method is comparatively easy to handle for most content

148. Sisk & Heise, *supra* note 2, at 792-93.

149. See generally ENCYCLOPEDIA OF SOCIAL MEASUREMENT (Kimberly Kempf-Leonard ed., 2005); KRIPPENDORFF, *supra* note 3; MOTIVATION AND PERSONALITY HANDBOOK OF THEMATIC CONTENT ANALYSIS (Charles P. Smith ed., 1992); NEUENDORF, *supra* note 26.

150. Wright & Huck, *supra* note 28; Hall et al., *supra* note 28.

analyses of judicial opinions. Most of the projects we reviewed studied all cases rather than a partial sample. Usually in social science, it is not feasible to observe all or most members of a relevant population. Therefore, the potential biases introduced by sampling method are a topic of considerable methodological focus and controversy. Various forms of random or representative selection are used to generate a study sample whose characteristics are likely to reflect the true population of interest—but how likely is often in doubt. Fortunately, most studies of legal decisions can avoid this raft of problems because the total sampling frame is small enough to manage.¹⁵¹ As Table 2 shows, 85% of case-coding projects used universal sampling limited only by year. Of these 114 universal samples, only 11 coded more than 1000 cases, and 21 coded from 500 to 1000. Twenty-six of these projects coded fewer than 100 cases (with 13 of these fewer than 51), and 39 coded between 100 and 300. All of these were universal samples restricted only by date, subject matter, jurisdiction, and/or source. In short, empirical researchers studying case law are usually able to avoid the selection bias issues that plague most other areas of social science.

When the total population is too large to be manageable, sampling techniques might include the following: (1) true random sampling, best done by a computer-generated list of random numbers; (2) systematic sampling, such as every fifth case; (3) quota sampling, such as all cases up to two hundred per jurisdiction per year; or (4) purposive sampling, such as cases that leading treatises, casebooks, or other cases cite. We observed examples of each of these methods.¹⁵²

For case analysis, more troubling than the sampling *method* is the potential for bias that arises from the sampling *frame*. Researchers who have done content analysis offer some practical advice about choosing a sample frame. Several recommend making a sample as homogenous as possible, both in subject matter and in procedural posture, if the goal is to analyze factors that predict case outcomes.¹⁵³ A more diverse selection of cases may be

151. See, e.g., Allison & Lemley, *supra* note 87 (noting that researchers avoided issues of sampling bias by coding the universe of all 299 relevant cases).

152. See, e.g., Benesh & Martinek, *supra* note 94, at 119 (stratified random sampling); Brace & Butler, *supra* note 24, at 247 (using both quota and stratified random sampling); Dixon & Gill, *supra* note 115, at 262-64 (random sampling); Falk, *supra* note 121, at 10 (random sampling); George & Sheehan et al., *supra* note 24, at 241 (random sampling); Johnson, *supra* note 122, at 181 (random sampling); Kagan et al., *supra* note 24, at 124-26 (systematic sampling); Koenig, *supra* note 32, at 1-3 – 1-5 (random sampling); Posner, *supra* note 11, at 34 (systematic sampling); Ross, *supra* note 111, at 506-07 (random sample from a cataloged collection); Schneider, *supra* note 94, at 332-33 (random sampling); Scott, *supra* note 27, at 1652-55 (random sampling); Siegelman & Donohue, *supra* note 41, at 1139, 1167 (quota sampling); Spriggs & Hansford, *supra* note 13, at 332-33 (random sampling); McChesney, *supra* note 48, at 134 (purposive sampling).

153. E.g., Allison & Lemley, *supra* note 87, at 201-05; Hammer & Sage, *supra* note 27, at 564; Songer & Haire, *supra* note 94, at 969.

appropriate, however, if the goal is to map broader trends and modes of judicial analysis.

Other researchers offer useful advice about which types of courts to select. One scholar recommends including trial decisions when mapping the landscape of opinions, but focusing instead on appellate decisions when analyzing factors that predict success or failure.¹⁵⁴ Others discuss choices between intermediate and highest appeals courts,¹⁵⁵ between state and federal courts,¹⁵⁶ and between published and unpublished opinions.¹⁵⁷

The choice of sample frame depends critically on the study's central questions and purposes. Analysts can narrow study questions to fit the available sample frame, or analysts can imagine a broader, theoretical universal of cases from which they draw a representative subset in order to pursue more interesting or important questions. Political scientists, for instance, often study political and institutional influences on judicial decision making by focusing on a particularly controversial or value-laden set of Supreme Court decisions.¹⁵⁸ Observing factors that influence free speech¹⁵⁹ or death penalty¹⁶⁰ cases, for instance, provides insights into judicial behavior more generally.

Too often, content analysts (including ourselves) choose sample frames without conscious reflection. One of us, for example, chose to explore how courts determine the effectiveness of medical treatment in health insurance disputes by studying all published judicial opinions resolving such disputes.¹⁶¹ That universe of observations might be relevant for a narrower question, such as how appellate courts explain their decisions regarding lower courts' rulings on such issues, but the sample frame of all published opinions does not fully reflect what all courts do or how they actually make their decisions because most court determinations are unpublished.

Whenever the actual cases selected do not fully match the sampling frame that theoretically applies to the questions posed or studied, sampling bias potentially exists. The findings from cases studied may not accurately represent those to which the authors apply their conclusions. For instance, studies that sample cases until a certain date cannot, necessarily, claim with confidence that their findings reflect what happened after that date.¹⁶² Similarly, studies that

154. Franklin, *supra* note 24, at 799-800 n.11.

155. See George & Sheehan, *supra* note 24, at 241; Songer & Haire, *supra* note 94, at 964.

156. See Brace & Butler, *supra* note 24, at 244-46; Kagan et al., *supra* note 24, at 121-22.

157. See Sisk et al., *supra* note 47, at 534-38.

158. Legal scholars have done this as well, looking at judicial review of agency decisions. See *supra* note 116.

159. See, e.g., Gruhl, *supra* note 107; Richards & Kritzer, *supra* note 129; Songer & Haire, *supra* note 94.

160. See, e.g., George & Epstein, *supra* note 94.

161. Hall et al., *supra* note 28.

162. See, e.g., Allison & Lemley, *supra* note 87, at 204-05 (noting that, even a universal sample of all relevant cases does not necessarily reflect the universe of future cases, and so may not accurately predict future results); Haar et al., *supra* note 7 at 660 (same). As a dramatic

sample from certain jurisdictions cannot claim their findings reflect other jurisdictions. Researchers should at least reflect on these potential distortions or limitations, and mention in their reports those that merit explanation.

Many case content analyzers do so, some at great length.¹⁶³ The most frequent and extensive reflections are researchers' agonized hand-wringing over whether available opinions and outcomes reflect all actual legal events of relevance. There is slippage at each point in the litigation process: most human interactions do not produce disputes, only some disputes result in legal claims, many claims are settled, and many trial decisions are not appealed. Appellate courts regularly dispose of cases without opinions or decide not to publish some opinions, and computer databases inconsistently include cases that are not officially published. At each juncture, a variety of factors potentially distort what one stage can reveal about the other.¹⁶⁴ These biases can fundamentally threaten the ability to generalize or the validity of a study's findings. In these situations, selection biases require thorough thought or, perhaps, major redesign of a study as originally conceived.

Other times, however, these considerations are not so worrisome.¹⁶⁵ No concern arises if the researcher defines the research questions in terms that match the population of cases actually sampled. For instance, unpublished

example, Mendelson, *supra* note 75, at 921-22, pointed out that Fred Kort's nearly perfect prediction of Supreme Court right-to-counsel cases up to the time of his publication would have ill-served a lawyer who later might have advised a convicted client against appealing based on Kort's analysis. See *supra* text accompanying notes 12-14. While Kort was doing his work, there was an unannounced shift underway in the Court's attitude towards these cases, resulting in defendants winning every single such appeal for the ten years leading up to *Gideon v. Wainwright* (which overtly changed the law). See LEWIS, *supra* note 12 at 128 (discussing this silent shift in the Court's right-to-counsel jurisprudence).

163. E.g., Allison & Lemley, *supra* note 87, at 203-05; Colker, *supra* note 111, at 244-47; Hammer & Sage, *supra* note 27, at 559-64; Richard L. Revesz, *Ideology, Collegiality, and the D.C. Circuit: A Reply to Chief Judge Harry T. Edwards*, 85 VA. L. REV. 805, 814-15 (1999); Schuck & Wang, *supra* note 27, at 126-28; Sisk et al., *supra* note 47, at 534-38. Some, however, completely ignore the issue, for instance, Absher & Baker, *supra* note 112, or misunderstand it, when it is fundamental to the questions they are analyzing. Professor Rice in several of his studies claims to test statistically for whether there is a selection bias by analyzing whether his findings differ significantly between lower court and appealed cases. E.g., Rice, *Judicial and Administrative Enforcement*, *supra* note 111; Rice, *Judicial Bias*, *supra* note 103. However, he is not able to test for whether reported or litigated cases differ from those that are settled or not reported.

164. For other examples surveying the literature and documenting various forms of case-selection bias, see Kevin M. Clermont & Theodore Eisenberg, *Litigation Realities*, 88 CORNELL L. REV. 119 (2002); Theodore Eisenberg & Stewart J. Schwab, *What Shapes Perceptions of the Federal Court System?*, 56 U. CHI. L. REV. 501, 502-04 (1989); Friedman, *supra* note 110, at 270-72; Marc Galanter, *Reading the Landscape of Disputes: What We Know and Don't Know (And Think We Know) About Our Allegedly Contentious and Litigious Society*, 31 UCLA L. REV. 4, 26 (1983); Henderson & Eisenberg, *supra* note 51; Daniel Kessler et al., *Explaining Deviations from the Fifty-Percent Rule: A Multimodal Approach to the Selection of Cases for Litigation*, 25 J. LEGAL STUD. 233 (1996); George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. LEGAL STUD. 1 (1984).

165. See, Trujillo, *supra* note 92, at 365-66.

opinions are irrelevant to the study of precedential law rather than the generalized behavior or attitudes of judges, so excluding them requires no justification. In other situations, where excluded cases are theoretically relevant, their exclusion can be justified easily by considering the likely direction of bias or distortion. When the bias runs in the same general direction as the study's findings—that is, the excluded cases are even more likely to exhibit the observed pattern—then including the additional cases would likely only strengthen the findings. The only harm from excluding them is potentially to have missed some additional findings of interest or to have falsely concluded that the relevant factor produced no effect. In situations where the expected differences between studied cases and omitted cases are sufficiently trivial, the omission should create no more a concern than other limitations that are obvious and inherent in the sample frame itself, such as using one date range rather than another.¹⁶⁶

All empirical studies are imperfect, especially observational (non-experimental) social science studies. The goal in selecting cases is not a perfect match between sample frame and research conclusions, but only a reasonable connection between the two. Inevitable imperfections in case selection methods often will not seriously threaten the entire validity of the study's findings. It usually suffices to acknowledge limitations fairly briefly.¹⁶⁷

In sum, as long as researchers do not extend their claimed findings beyond their studied universe, they should not agonize because their chosen sample frame is not capable of conclusively answering all conceivably relevant questions—especially when they have selected all cases within their defined frame. Only when researchers make strong claims about how their findings apply beyond the sampling frame, to a larger population that is probably different than the one they studied, do their publications need to consider at length these potential selection bias and validity problems.

2. Selection Techniques and Replicability

An essential attribute of scientific objectivity is *replicability*—the ability, at least in theory, to test a research project by reproducing its findings using the same methods.¹⁶⁸ Replicability is the overriding reason for using systematic

166. In this article, for instance, we sample only *published* empirical studies that use case content analysis. We therefore cannot make claims about researchers who fail to complete or publish their content analyses. For our purposes, however, that limitation is not important, or is obvious, even though it excludes some questions of potential interest, such as whether unobserved techniques are more difficult to complete or to publish. In a similar vein, we think some researchers and critics have made much too big a deal about the fact that case content analysis omits non-decided or unpublished cases.

167. See, e.g., Juliano & Schwab, *supra* note 76, at 556-57 (offering a nice example that other legal scholars might emulate, of an efficient—single short paragraph—acknowledgement of the inherent limitations of analyzing only a certain set of available opinions).

168. See Epstein & King, *supra* note 2, at 38, 83 (stressing the scientific importance of

content analysis. This is what confers scientific status on research findings.

One component of replication is the method of selecting cases for study. On this score, case content analysts are doing well. The great majority of projects we reviewed selected cases in a fashion that is at least partially replicable.¹⁶⁹ Most case coding projects in the past fifteen years have taken advantage of the ability to select cases using structured computer searches in the Westlaw or LEXIS databases. Researchers usually specify the exact search terms and strategies so that others can replicate their methods, but many studies still fall short of a fully replicable case selection because computerized searches are rarely refined enough to narrow the sample to relevant cases.¹⁷⁰ Supplemental techniques are needed to screen out cases that mention a topic of interest only in passing or those that decide an issue on technical or procedural grounds irrelevant to the study. Most legal researchers do so using subjective criteria of relevance that cannot be fully replicated. Researchers could instead refine their mechanical search strategy rather than screening initial selections one by one. Useful strategies include searching case digests or headnotes rather than the full case itself, or searching a sample of cases that cite particular statutes or that appear in a subject matter classification drawn by someone else, such as West's key numbering system or the publisher of a subject-matter-specific reporter. In effect, such researchers are relying on case selection criteria employed by someone else to establish probable relevance of cases.¹⁷¹

We found no study that formally tested the replicability (or statistical reliability) of its case selection. Even researchers (such as us) who used state-of-the-art methods to test the reliability of their case *coding* did not test their case *selection* methods. Replicating case selection would make a study more scientifically rigorous by eliminating the possibility that a researcher subconsciously chose cases according to whether they appear to support the researcher's preliminary hunches. Methods to guard against this potential bias include formal reliability testing or having someone who is otherwise uninvolved in the study select the cases. Failing to do this, however, is often not a major flaw because, in our judgment, case selection criteria are usually reasonably objective and so would likely produce a similar selection if other researchers were to employ them.

replicability).

169. We coded at least partial replicability for 87% of projects (117/134), but our intercoder agreement for this measure was not reliable, so we present this observation only as a qualitative assessment. See online Appendix, Hall & Wright, *supra* note 29

170. In this context, full replication is not achieved simply by listing all case citations or providing the researcher's database. Cf. Morriss, *supra* note 47, at 1096-1142 (providing such a list). Here, replication refers to *how* cases were selected rather than know *which* cases were in fact included. A pure random sample is easily replicated as a selection *method* even if it does not produce precisely the same list of cases. In contrast, an exact listing of case citations that the research happened to find interesting following an ad hoc search is entirely nonreplicable, even though we know exactly what was in the sample.

171. Discussing this point, see Phelps & Gates, *supra* note 122, at 584-88.

B. Coding Cases

Once cases are selected, experts in content analysis outline four basic steps researchers should follow in coding any material¹⁷²:

- (1) Based on the questions that seem most germane to the study, create a tentative set of coding categories a priori. Refine these categories after thorough evaluation, including feedback from colleagues, study team members, or expert consultants.
- (2) Write a coding sheet and set of coding instructions (called a "codebook"), and train coders to apply these to a sample of the material to be coded. Pilot test the reliability (consistency) among coders by having multiple people code some of the material.
- (3) Add, delete, or revise coding categories based on this pilot experience, and repeat reliability testing and coder training as required.
- (4) When the codebook is finalized, apply it to all the material. Then, or during that process, conduct a final, formal reliability test.

This Section elaborates on each of these steps.

1. Coding Categories and Instructions

The categories used to code the content of judicial decisions are tremendously diverse, owing to the wide range of questions that researchers pursue. Commonly used factors might be sorted into four general groups: (1) the parties' identities and attributes, (2) the types of legal issues raised and in what circumstances, (3) the basic outcome of the case or issue, and (4) the bases for decision. Coders often do not distinguish the "facts" of a case from various arguments that are made. Instead, they usually code simply for whether a variety of factual or legal factors are present in the case in some fashion.¹⁷³ Coders should consider whether it suffices if these factors are merely alleged, realizing that the allegations may be sharply contested. If mere allegations are not sufficient, what is? Obviously, the procedural posture of a case (summary judgment versus post-trial) can complicate this evaluation.

In coding the bases of decisions, coders frequently distinguish between procedural and substantive rulings and they record the various types of authorities that courts cite or rely upon. Some researchers also code for the degree of importance that various factual or legal factors have in the court's analysis or holding.¹⁷⁴ Coding also commonly focuses on the court's style of

172. KRIPPENDORF, *supra* note 3, at 129-30; NEUENDORF, *supra* note 26 at 50-51. A summary flowchart is available at <http://academic.csuohio.edu/kneuendorf/content/resources/flowc.htm>.

173. *E.g.*, Absher & Baker, *supra* note 112; Ignagni, *supra* note 140; Johnston, *supra* note 27; Kritzer & Richards, *supra* note 96; Ostas, *supra* note 58; Posner, *supra* note 11.

174. *E.g.*, Brudney & Ditslear, *supra* note 76; Henderson & Eisenberg, *supra* note 51; Jacobson et al., *supra* note 76; Schuster, *supra* note 111.

analysis or approach to statutory or constitutional interpretation.¹⁷⁵

Coding is not restricted to manifest variables that are explicit in the text; it has been shown to work well also for some “latent” variables that require inference or evaluative judgment.¹⁷⁶ For instance, Charles Johnson demonstrates law students’ ability to code cases with some degree of reliability for the clarity, complexity, and completeness of their discussion of facts, issues, holding, reasoning, and the law.¹⁷⁷

Coding experts advise researchers to create more coding categories, and to make coding more fine-grained, than they may ultimately want to analyze. Though this produces more information than the project will eventually require, this process enables the researcher to test different categorization schemes to learn through trial and error which work best. Ultimately, the goal is to maximize the exhaustiveness of coding using mutually exclusive categories—in other words, to capture all the relevant information, but to avoid having categories that duplicate or overlap each other.¹⁷⁸ This does not mean, however, that the coding must reflect each possible nuance. Instead, categories should be used only if they occur frequently enough in the material to merit attention. Rare or unusual factors can be coded simply with a miscellaneous “other” option.

Categorizing case outcomes provides a good example of exhaustive and mutually exclusive coding. Defining what counts as a win or loss across a range of cases is not a simple matter. Appellate cases arise in a variety of procedural postures, involve multiple issues, and each issue can be resolved in several different ways. Case coding projects often have to devise complex categories to capture all the relevant detail. The U.S. Court of Appeals Database, for instance, defines all possible case outcomes using nine categories.¹⁷⁹ This illustrates that it is a better practice to be over-inclusive at the coding stage, waiting until the analysis stage to collapse the various categories into discrete

175. E.g., Brudney & Ditslear, *supra* note 76; Czarnezki & Ford, *supra* note 95; Daynard, *supra* note 21; Phelps & Gates, *supra* note 122; Rebell & Block, *supra* note 11; Schacter, *supra* note 27; Schneider, *supra* note 94.

176. See NEUENDORF, *supra* note 26, at 23 (reviewing several studies that used latent constructs as a way of integrating content analysis); Little, *supra* note 50, at 91-93 (coding Supreme Court cases for the degree of linguistic obfuscation that the coder felt was present in the opinion).

177. Johnson, *supra* note 122, at 182-96. Some researchers have clearly stretched this too far, however. For instance, Falk, *supra* note 121 at 13 n.47, used a 7-point Likert scale to score opinions for their degree of reliance on social science evidence, and Groscup et al., *supra* note 115 at 343, had students use a 10-point scale to rate the degree of influence various factors had on a decision. Still, it is possible, as described below, to aggregate fine-grained coding to a coarser resolution to obtain reasonable reliability. See *infra* text accompanying note 219.

178. Reed C. Lawlor, *Fact Content of Cases and Precedent - A Modern Theory of Precedent*, 12 JURIMETRICS J. 245, 264 (1971-72).

179. George & Sheehan, *supra* note 24, at 241. See also, Lee Epstein & Andres Martin, *Coding Variables*, in ENCYCLOPEDIA OF SOCIAL MEASUREMENT 323 (Kimberly Kempf-Leonard ed., 2005) (discussing this coding variable in particular).

win/loss columns.

When categories are finalized, good coding practice requires researchers to record their description of the working categories and their instructions for applying the categories. This is necessary not only if someone other than the researcher, such as student assistants, performs the coding; it is also necessary even if authors do their own coding because the scientific standard of replicability requires a written record of how categories were defined and applied. In our literature review, we found that researchers in 60% (80/134) of projects created a codebook, whether or not one is actually available now.¹⁸⁰ As an aid to future researchers, and to promote scientific replicability, the *Wake Forest Law Review* plans to provide a repository for codebooks, coding sheets, and datasets used in content analyses of judicial opinions.¹⁸¹

Experienced coders advise that designing coding forms to minimize writing will reduce errors. For instance, a form might provide a checklist of factors to indicate presence or absence by ticking boxes rather than having to write in a number or letter.¹⁸² However, researchers should strike a balance between an overly spare form that offers almost no instructional information for coders, forcing them to refer frequently to the detailed coding manual, versus an overlong coding form that contains a full set of instructions.¹⁸³ One legal researcher reports, based on extensive testing of student coders, that it does not typically help to extensively revise succinct, well-written coding categories simply “to satisfy the whim” of each coder who might ask for more detailed instructions.¹⁸⁴ It is inevitable that some measure of ambiguity will remain in how coding categories should apply to particular cases. Often, there is no obvious right way to resolve these judgment calls but such ambiguity is not disabling as long as coders are reasonably consistent in how they apply coding categories across a range of cases.

2. Choosing and Training Coders

A major dilemma in coding cases is whether principal investigators should do this work themselves, or instead whether they should supervise students (or others). Only 22% (29/134) of the projects we reviewed primarily used student coders; in the rest, the authors appeared to do their own coding.¹⁸⁵ In theory,

180. Sometimes the publication itself was sufficiently detailed that its text might be used as a codebook, or the codebook and coding form were included as an appendix.

181. Available at <http://www.law.wfu.edu/x5649.xml>. Researchers who wish to contribute their materials should contact the Editor-in-Chief, whose contact information is available at <http://www.law.wfu.edu/lawreview.xml>.

182. See, e.g., KRIPPENDORF, *supra* note 3, at 148.

183. For good examples of well-designed code books and coding forms, see Colker, *supra* note 111; Hammer & Sage, *supra* note 27; Lawlor, *supra* note 134; the research web pages linked to <http://epstein.law.northwestern.edu/cv.php> (last visited July 15, 2007).

184. Lawlor, *supra* note 134, at 129.

185. Sometimes, this was not entirely clear, but unless there was clear evidence of coding

the most scientifically rigorous method is for researchers to train others to do the coding and for coders to work completely independently once they are trained.¹⁸⁶ (Note the similarity to how law professors teach and then examine law students.) Using generic coders ensures that the researchers' preliminary hypotheses and personal views do not bias the coding too much.¹⁸⁷ Also, this can save researchers considerable time and effort in large coding projects. Moreover, the imposed discipline of training and supervising coders ensures that coding instructions are written in a way that others can follow.¹⁸⁸ These attributes promote the reproducibility that is essential for good science.

Coding by law students is appropriate when some general legal knowledge is required but it is not necessary to be an expert in the field of study. Although some students inevitably will be more diligent than others, one researcher reports that the skill of student coders does not necessarily depend on their class rank.¹⁸⁹ Rather, this researcher found that coding reliability improved the more that students were trained.¹⁹⁰ Researchers who use students should describe how training was done in sufficient detail such that others could replicate all essential steps.¹⁹¹

Other considerations might counsel against using student coders. Training coders to achieve accurate and reliable results can be difficult and time-consuming. This may require considerably more resources and effort than researchers doing their own coding, especially in smaller projects.¹⁹² A relevant

by students or others, we assumed the authors did their own coding. In some instances, the authors only resolved questionable cases or corrected errors, which we still regarded as coded primarily by students. If students assisted authors but the authors made their own final decisions for most of the coding, we counted this as author coding. We also counted as author coding the projects where students were themselves an author and did their own coding. The key criterion is not the coder's level of expertise but instead whether the coder also conceived, designed, and conducted the overall study.

186. Experts caution against allowing coders to confer with each other once they are trained because this undermines the independence of their judgment, thus calling into question the objectivity and reproducibility of their coding. Johnson, *supra* note 122, at 175.

187. In fact, one expert recommends that researchers not even tell coders the study's purpose and main hypotheses. See NEUENDORF, *supra* note 26, at 133. We are not aware, however, that this advice is frequently followed.

188. According to KRIPPENDORFF, *supra* note 3, at 131, "[s]elf-applied recording instructions are notoriously unreliable."

189. Lawlor, *supra* note 134, at 119.

190. One expert cautions, however, against excessive training of coders to avoid instilling in them the researcher's personal views in a fashion that cannot be replicated. See KRIPPENDORF, *supra* note 3, at 130-31.

191. Good examples are Hammer & Sage, *supra* note 27, at 554 and Johnson, *supra* note 122, at 181-82.

192. Several researchers emphasize how much more difficult this undertaking proved to be than they first anticipated. *E.g.*, Bliss Cartwright, *supra* note 144, at 376 n.10 ("coding is not cheap. Good coders are expensive and hard to find. Error-checks are absolutely essential, tedious and time-consuming. Never underestimate the number of errors from even the best trained coders."); Haar et al., *supra* note 7, at 656; Kagan et al., *supra* note 24, at 128 (coding was "an immense, almost intractable task"); Lawlor, *supra* note 134, at 116-22 (describing extensive

selection of cases is often sufficiently small that a single reader can handle the coding alone. In our review, twenty-nine projects (22%) coded fewer than one hundred opinions and another twenty-eight (for a total of 42%) coded fewer than two hundred.

Several researchers emphasized the surprising number of mistakes student coders make, even on seemingly simple and objective criteria such as dates.¹⁹³ We observed several instances of researchers who first began, but then abandoned or altered, student coding because of the difficulties and frustrations.¹⁹⁴ Although using student coders may promote reliability, this can threaten the validity of results if the information students record is not accurate or is too simplified to be very meaningful.¹⁹⁵ Student coders may lack the level of expertise needed to code reliably the more complex or subtle, yet more meaningful, aspects of judicial opinions. If so, researchers will be sorely tempted to do their own coding, and as noted above, most have. When this is done, however, it is especially important from a scientific perspective to conduct reliability tests, by recruiting a colleague with similar expertise to double code at least a subset of cases.¹⁹⁶

The dilemma over who should code goes to the core of the epistemological differences between traditional and empirical methods of legal scholarship. Traditionally, the scholars with the most expertise are best able to interpret cases. When each case must receive an individualized reading to develop the scholarly thesis, the scholar does the reading and interpretation for herself, with students helping only to locate the relevant cases. From a social science perspective, however, this way is the height of unmitigated subjectivism—the opposite of good scholarship.¹⁹⁷ Scientifically, scholarship

efforts to train coders, which improved their performance but never achieved perfection); Phelps & Gates, *supra* note 122, at 588-90 (describing extensive pre-testing, which took 3 months, and coding that took 18 months).

193. We add the qualifier “seemingly” because even dating events at the level of a year is sometimes not self-evident. In our own review of content analysis projects, our coding for the year did not agree 12.5% (4/32) of the time, because of uncertainty about when the project’s methods and findings were first published, or due to simple error. Morriss, *supra* note 47, wrote a one hundred fifty-page article developing and testing eighteen general principles about how to date changes in the law announced through judicial decisions.

194. *E.g.*, Colker, *supra* note 111; Winkler, *supra* note 111.

195. See KRIPPENDORF, *supra* note 3, at 212-13 (discussing the trade-off between reliability and validity).

196. In our literature review, this was done in only five of the ninety projects where authors did some or all of their own coding. See Jacobson et al., *supra* note 76; Phelps & Gates, *supra* note 122; Songer & Haire, *supra* note 94; Spaeth, *supra* note 24; James F. Spriggs & Paul J. Wahlbeck, *Amicus Curiae and the Role of Information at the Supreme Court*, 50 POL. RES. Q. 365 (1997).

197. One of us (Hall) came to this realization the hard way, after repeated failures to obtain NIH funding for his study of health insurance disputes. Hall et al., *supra* note 28. His first three tries proposed conventional legal research methods. He was successful on his fourth try only after converting his methods to content analysis using student coders. He concludes from this experience that, if a tenured full professor of law proposes to analyze case law, NIH scientific

done this way reflects only what one person happens to think a legal opinion means—just a single datum with almost no empirical significance. Law professors who code their own cases, without any reliability checks or coding instructions, are trying to keep one foot on each side of this epistemological divide. Legal scholars must decide which mode of interpreting and analyzing cases best suits their purposes. Otherwise, they risk producing work that both camps will judge to be incompetent.

3. *Testing Reliability*

As we have stressed, demonstrating the reliability of coding is an essential aspect of good content analysis. “[G]iven that a goal of content analysis is to identify and record relatively objective (or at least intersubjective) characteristics of messages, reliability is paramount.”¹⁹⁸ If coding categories are so objective and straightforward that consistent application is clearly likely, then perhaps this step is not necessary. If there might be elements of subjectivity or uncertainty in applying coding categories to legal decisions, any claim to scientific rigor requires some evaluation of whether different people would code the documents consistently. This is essential because the theory of coding—the reason systematic content analysis is done at all—is the implicit claim of reproducibility, that other researchers using the same methods will achieve approximately the same results. Coding that primarily reflects the subjective, idiosyncratic interpretation of the particular individuals who read the cases or that has large elements of error or arbitrariness undermines the claim of replicability.

Nevertheless, formal statistical reliability testing is not mandatory. Even without any reliability testing, it is possible that a coding scheme in fact is reliable; as long as the researcher describes the coding in sufficient detail for others to replicate, she can choose to let others challenge or verify reliability.¹⁹⁹ Indeed, 65% (87/134) of projects we reviewed had no discussion of coding reliability whatsoever, and 21% (28/134) discussed coding reliability in only general qualitative terms, leaving only 14% (19/134) of projects with any type of quantitative assessment of coding reliability.²⁰⁰ Considering the inherent

reviewers see this as little more than impressionistic literary interpretation, but if the professor hires three law students to do the same thing, suddenly it becomes rigorous social science.

198. NEUENDORF, *supra* note 26, at 141. See also Matthew Lombard, Jennifer Snyder-Duch & Cheryl Campanella Bracken, *Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability*, 28 HUMAN COMMUNICATION RES. 587 (2002).

199. See, e.g., Spriggs & Hansford, *supra* note 13 (establishing the reliability of coding done by Shepard’s citation service). This is also the approach taken in Dnes & Seaton, *supra* note 129, at n.24 (inviting readers to check the authors’ work), but in this instance, we think the factors are obviously too subjective to meet social science standards of replicability. They include, in a study of manslaughter and fatal accident cases, the ease of detecting the event, which of two parties is the least-cost risk avoider, and whether there is a non-linear relationship between risk of death and compensation for risk.

200. Surprisingly, these percentages were not dramatically different within the subset of

subjectivity of reading cases, legal empiricists should do better about testing reliability.

Good technique requires content analysts to evaluate coding reliability in some fashion.²⁰¹ The best method is to conduct formal reliability tests during at least two stages in the process: initially, while piloting the draft coding instructions, and later, once coding categories and instructions are completed.²⁰² Formal testing calls for at least two coders independently to code a sample of cases and to compare their results statistically.²⁰³ The most common statistic is simple percent of agreement, but a simple percentage does not account for the level of agreement that would be expected purely by chance.²⁰⁴ Because chance agreement varies according to the type of coding scheme (for example, a variable with two possible answers will naturally produce more agreement than a variable with eight possible answers), the best practice is to report one of several coefficients that reflect the extent of agreement beyond what is expected by chance. There are several such statistical tests, the most common of which is known as “Cohen’s kappa,” named after its inventor, or simply the *kappa* statistic.²⁰⁵ Ranging from 0 to 1, *kappa* indicates the proportion of observed agreement that exceeds what would

projects conducted by social scientists; 58% (29/51) had no reliability discussion, and only 25% (13/51) reported formal, quantitative tests. Reviews in other social science literatures document not-too-dissimilar practices in other research fields. See Lombard et al., *supra* note 198, at 599; NEUENDORF, *supra* note 26, at 142.

201. Model discussions of reliability testing can be found, for instance, in Gerard F. Anderson, Mark A. Hall, & Teresa R. Smith, *When Courts Review Medical Appropriateness*, 36 MEDICAL CARE 1295, 1297 (1998); Juliano & Schwab, *supra* note 76, at 598-602; Richards & Kritzer, *supra* note 129, at 311-12, 312 n.8; Lawlor, *supra* note 134, at 129.

202. Demonstrating the value of multiple layers of training, Lawlor, who reports on extensive testing of coding methods by law students, found that reliability among a group of nine students averaged 82% after only a brief orientation and training session, but it ranged from 73% to 90% across students. Lawlor, *supra* note 134, at 116-22. After extensive training, reliability improved to an average of 95%, ranging much more tightly from 94% to 97% across students. *Id.*

203. There is no standard rule for how many cases should be double-coded, but a recommended rule of thumb is at least 10% of the sample or thirty, whichever is less. Lombard et al., *supra* note 198; NEUENDORF, *supra* note 26, at 158-59. For our coding of content analysis databases, we double-coded thirty-two of the one hundred thirty-four databases (24%).

204. For instance, if there are two coding options, such as win or lose, and they occur evenly in the sample cases, two coders can be expected to agree purely by chance (as if they coded without even reading the cases) one-half of the time. NEUENDORF, *supra* note 26 at 150.

205. An older statistic is Scott’s *pi*, and a newer one is Krippendorff’s *alpha*. Some statistics apply only to simpler coding schemes, such as two options coded by two people, whereas others can be used for multiple coders applying more complex categories. Statisticians debate the relative merits and flaws of these coefficients, but their use is also affected by practical considerations such as the difficulty of calculating by hand and whether they are available in commonly used statistical software. One reason that use of Cohen’s *kappa* is so widespread is that it is included in the Window’s version of SPSS, and free versions of “macro” add-ons to SAS are available. For more information, including instructions for calculation and access to software, see MATTHEW LOMBARD, JENNIFER SNYDER-DUCH, CHERYL CAMPANELLA BRACKEN, PRACTICAL RESOURCES FOR ASSESSING AND REPORTING INTERCODER RELIABILITY IN CONTENT ANALYSIS RESEARCH PROJECTS (2005), <http://www.temple.edu/mmc/reliability>.

be expected by chance alone, with 0 indicating agreement entirely by chance and 1 indicating perfect agreement.²⁰⁶

If researchers use these statistics, they must do so correctly. One mistake is to test the overall reliability of all variables combined. The correct method is to test each variable's reliability because reliability can vary widely, so aggregate statistics can mask serious problems with key variables.²⁰⁷ Also, when the response pattern for a variable is highly skewed (meaning that one or some available responses occur much more frequently than the others), this should be noted or taken into account; otherwise, the nominal level of agreement can be deceptive. Haar and colleagues, for instance, coded for presence or absence of one hundred and sixty-seven different factors in each case.²⁰⁸ Most likely, only a dozen or so appeared in any one case. Haar did not test for coding reliability, but if he had, he most likely would have found a very high percent of agreement simply because most factors would not be present in most cases. The key question, instead, is whether coders agree when they indicate a factor is present.²⁰⁹

For content analysts who do not have access to statistical testing, another alternative is to report the simple percent of agreement for the group of observations where the less frequent variable appears. For instance, if coders are reviewing cases for a long list of factors and "not present" is far and away the most frequent response, instead of reporting agreement overall, reliability can be reported separately for the subset of cases where one or both coders indicated the factor was present.²¹⁰ Coders might be surprised by how low the nominal percent of agreement turns out to be. In our review of content analysis publications, for instance, we coded for whether or not authors reported statistical measures of reliability. In our reliability testing of our own coding, we agreed 75% of the time overall for this factor, but only 25% of the time in the 12.5% of publications for which one or the other of us indicated that a statistical test was used, producing a *kappa* statistic of 0.43.²¹¹ Most of our

206. See NEUENDORF, *supra* note 26, at 150; Lombard et al., *supra* note 198, at 592. A negative *kappa* statistic indicates less agreement than expected by chance.

207. Lawlor, *supra* note 134, at 124-25. In our coding, the raw percentage of agreement across ten key variables ranged from a high of 100% to a low 37.5%, with *kappa* statistics ranging from 1 to 0.06. See online Appendix, Hall & Wright *supra* note 29.

208. Haar et al., *supra* note 7, at 60.

209. Neuendorf demonstrates, for instance, that if responses to a bivariate coding category split 90/10, then the odds of agreeing totally by chance are 82%; therefore, even 90% agreement overall would have a reliability coefficient of only 0.44. See NEUENDORF, *supra* note 26, at 151. For additional discussion and illustration of this point, see Johnson, *supra* note 122, at 183; Spaeth, *supra* note 24, at xviii.

210. For an example of this approach used in the review of medical records, see A. Russell Localio, et al., *Identifying Adverse Events Caused by Medical Care: Degree of Physician Agreement in a Retrospective Chart Review*, 125 ANN. INTERN. MED. 457, 460 (1996) (finding only 44% agreement in the 5% of records reviewed where one or both physicians coded an "adverse event").

211. Similar to Lawlor, *supra* note 134, at 124, our discrepancies in this and other

agreement happened in the cases where we both found that no reliability testing was reported (which, fortunately, is the more important category for us).

When reliability testing reveals discrepancies, as it almost always will, this will usually reveal unresolved questions in the coding instructions. The analyst can correct these problems if the error appears after the pilot phase rather than after the completed coding. Compulsive legal researchers might try to get to the bottom of remaining disagreements and resolve all discrepancies before beginning final coding. This is unnecessary and impractical. Disagreements sometimes arise from overt errors, but often they result simply from judgment calls or inevitable ambiguities that may be virtually impossible to eliminate without compromising the independence of individual coders.²¹²

Refining coding rules to eliminate all elements of ambiguity is usually not possible, no matter how prescriptive the rules. Plus, each time the rules are rewritten, the researcher should retest the refined rules for reliability. This can produce a never-ending cycle in search of elusive perfection.²¹³ In short, coders should learn to live with a certain degree of imperfection once coding is found to be reasonably reliable, and draw appropriately modest conclusions when relying on variables with weaker levels of inter-coder reliability.²¹⁴

Although there is broad agreement on the desirability of testing for reliability and some agreement on the methods for doing so, there is no firm agreement on what level of reliability is acceptable. The goal is aspirational—to achieve high levels of agreement rather than merely to rise above purely random agreement. One expert says that a reliability coefficient of 0.80 (that is, data agree 80% more than mere chance agreement) is good, with indices from 0.67-0.80 being sufficient for “tentative conclusions.”²¹⁵ Others claim that this is too demanding,²¹⁶ especially for coding categories that produce more skewed responses, since even small levels of disagreement can cause the statistical

categories were due to a mix of simple errors, incomplete coding instructions, and inherently ambiguous or judgmental coding categories.

212. See *id.*

213. Lawlor, *supra* note 134, at 121-23, is a sobering example of a researcher's futile attempt to achieve perfect reliability among law student coders. Speaking tongue-in-cheek, the researcher described part of the process as follows: “We thereupon placed the six analysts in a closed cell (called a seminar room) and vibrated them at high frequency under conditions of high pressure and high temperature. The debates were vigorous. Sometimes the air was saturated with thunder and occasionally even piercing acrimony.” *Id.* The process, which spread over many three-hour sessions and lasted a total of 25 hours, improved agreement to 99%, but perfection was never reached. *Id.*

214. Still, when reliability testing identifies disagreements, some decision is needed about which of the two codes to use. For this, too, there is no hard-and-fast rule. If overall coding is reliable, in theory it is acceptable to use either of the conflicting codes since by definition they are equally reliable, but most researchers will assert their own expertise to resolve the disagreement, or resort to another tie-breaking technique such as coding by a third person. See Lombard et al., *supra* note 198.

215. See KRIPPENDORF, *supra* note 3, at 241.

216. See Lombard et al., *supra* note 198, at 593.

index to drop rapidly.²¹⁷ Therefore, other authors provide a more lenient classification for the *kappa* statistic: <0.00 is Poor; 0.00-0.20 is Slight Agreement; 0.21-0.40 is Fair; 0.41-0.60 is Moderate; 0.61-0.80 is Substantial; and above 0.80 is Almost Perfect, since, technically, 1.00 is perfect.²¹⁸ Keep in mind that these recommendations are for agreement levels beyond what is expected by chance. For raw, unadjusted percents, agreement levels below 70 or 80% are usually not considered to be good.

If coder agreement is not acceptable, researchers must retrain coders, revise their coding categories, decide not to use the data, or use the data but with appropriate caveats. Following best practices, the first two options call for retesting of reliability. One convenient remedy is to combine marginally reliable detailed coding into a more aggregated category that has good reliability. Drawing again from our own coding of content analysis publications, we had poor agreement, 37% (12/32), for whether case selection was fully, partially, or not at all replicable, but after combining the first two categories, we had adequate agreement, 81% (26/32), for whether case selection was replicable to any extent.²¹⁹

4. *Alternative Coding Techniques*

Researchers might consider alternatives to independent coding by students or experts. Sometimes a *group* of students code each case and the researcher assigns the outcome coded by the majority,²²⁰ a technique that mimics the deliberative structure of an appellate court panel. Group coding creates the impression of greater objectivity, and may in fact improve reliability, but this is not necessarily the case. Resolving split votes with whatever the third person thinks might be as arbitrary as using a single coder.²²¹ The only way to find out for sure is to test the reliability of panel coding by coding a sample of cases independently with a different panel.

Similarly, some researchers ask coders who initially disagree to confer and seek consensus, or the researcher uses her own expertise to resolve disagreements. Again, this may or may not improve reliability, but it does not *establish* reliability. The process of reaching consensus might be arbitrary,²²² or the author's expert view may not be objectively reproducible.

A variation of these techniques is an expert panel consensus model, which

217. See *supra* text accompanying notes 210-11.

218. J. Richard Landis & Gary G. Koch, *The Measurement of Observer Agreements for Categorical Data*, 33 *BIOMETRICS* 159, 165 (1977).

219. See also, Johnson, *supra* note 122, at 184-85 (comparing agreement rates for two-category versus four-category coding).

220. Phelps & Gates, *supra* note 122, at 587; Johnson, *supra* note 94.

221. See KRIPPENDORF, *supra* note 3, at 217.

222. Lawlor, *supra* note 134, at 121-22, for instance, reports that his student coders reached consensus sometimes "by majority decision after vigorous debate, some by quick recognition of obvious errors, some by acquiescence in interpretation by others, and some even by log rolling."

was developed for evaluating medical judgments. This has not been used so far for legal judgments but it is worth exploring. Known as the Delphi technique, each expert first rates a case independently, then learns how peer experts have rated it, and then, following discussion, each expert gives an independent final rating, with the majority controlling when there is not unanimity.²²³ This has been shown to be a fairly reliable method for rating highly complex and judgmental aspects of medical decision making.²²⁴ It combines elements of “gold standard” expertise with consensus building and majority rule.

Finally, there is an innovative technique that avoids the vagaries altogether of training coders and demonstrating reliability: using completely mechanical forms of content analysis that can be done by computer or simple computation. For instance, some studies count the number of words or paragraphs devoted to discussing particular factors as an indication of the factors’ relative importance.²²⁵ Also noteworthy are researchers who analyze judicial texts entirely by computer, looking for revealing patterns in syntax or semantics.²²⁶ For example, one study compared the linguistics of draft opinions circulated among Supreme Court chambers with the final versions to gauge the relative influence that clerks have on different Justices.²²⁷

C. Statistical Analysis

A credible content analyst does not always need to use complex or sophisticated statistics—or, indeed, any statistics at all. In our literature review, 29% (39/134) of projects (and 34% (27/80) of those published by lawyers) performed no statistical testing of any kind. Instead, they relied on counts and frequencies (percents) to show how often a given feature appears in the cases. Quantitative descriptive analysis such as this has empirical value, in the same way that a naturalist might report how many of various species were found in a

223. See Rolla Edward Park, et al., *Physician Ratings of Appropriate Indications for Six Medical and Surgical Procedures*, 76 AM. J. PUB. HEALTH 766 (1986) (applying “Delphi technique” to generate consensus among expert panel of physicians regarding appropriate use of medical procedures).

224. See Paul G. Shekelle, et al., *The Reproducibility of a Method to Identify the Overuse and Underuse of Medical Procedures*, 338 NEW ENG. J. MED. 1888 (1998) (evaluating the reliability of the Delphi expert consensus technique).

225. See, e.g., Groscup et al., *supra* note 115; Wahlbeck, *supra* note 94.

226. See, e.g., Kevin T. McGuire & Georg Vanberg, *Mapping the Policies of the U.S. Supreme Court: Data, Opinions, and Constitutional Law* (Prepared for delivery at the Annual Meeting of the Am. Political Science Association, Sept. 2005), available at http://www.unc.edu/~kmcguire/papers/McGuire_and_Vanberg_2005_APSA_Paper.pdf (demonstrating the “wordscore” technique of analyzing two ideologically polarized decisions to derive an algorithm for scoring policy positions based on key words used); Michael C. Evans, et al, *Recounting the Courts? Applying Automated Content Analysis to Enhance Empirical Legal Research*, 1st Annual Conference on Empirical Legal Studies Paper (August 28, 2006), available at <http://ssrn.com/abstract=914126> (similar).

227. See Paul J. Wahlbeck, James F. Spriggs, & Lee Sigelman, *Ghostwriters on the Court? A Stylistic Analysis of U.S. Supreme Court Opinion Drafts*, 30 AM. POLITICS RES. 166 (2002).

geographic exploration. Often, simple counts and percents are sufficient to document more rigorously an author's claim about trends in the case law, to challenge conventional wisdom, or to raise provocative questions meriting further study. Because case-counting studies often code the entire universe of relevant cases,²²⁸ statistics are not essential for analyzing the probability that the sample cases reflect the reality in a larger population.²²⁹ Moreover, content analysis need not involve numbers at all. Instead, it can employ rigorous methods of purely qualitative analysis²³⁰ that focus on themes and patterns that are best described through conceptual description and narrative illustrations rather than numbers.²³¹

Still, statistical testing often helps to make the most use of a data set. This is not the place to outline the full range of statistical approaches, but we will note two broad types of statistical analysis: bivariate and multiple regression. Bivariate analysis tests whether one count differs significantly from another, or instead whether the difference might be due entirely to chance. This is useful, for instance, when noting that a particular factor occurs more often in one set of cases than in another set. Thirty percent (28/95) of the publications that used statistical analysis used only bivariate statistics (38% (20/53) among legal researchers).

One danger in using statistical testing in exploratory studies is that, without a tightly controlled analytical focus, such as a predefined set of hypotheses that are being tested, it becomes too easy to find associations and patterns of apparent significance entirely by chance. Statisticians denigrate this approach as a "fishing expedition."²³² If enough variables are examined and enough comparisons are made, odds are that significant findings will emerge, but some or all of these apparent findings could be due entirely to chance without additional statistical adjustments for the number of possibilities that were explored.²³³

228. See *supra* Table 2.

229. Testing for statistical significance does help, however, to indicate the likely "margin of error" in observations due to chance variation in when observations were made, how cases arose, and which cases were selected.

230. Often, researchers use computer programs such as Nudist and Ethnograph to assist with this analysis. See *generally* COLLECTING AND INTERPRETING QUALITATIVE MATERIALS (Norman K. Denzin & Yvonna S. Lincoln eds., 2003); M.Q. PATTON, QUALITATIVE RESEARCH AND EVALUATION METHODS (3rd ed. 2002).

231. Most of traditional legal scholarship is qualitative in some sense, but does not use the tools and techniques that social scientists would regard as scientifically rigorous. A notable exception employing case content analysis is Fradella et al., *supra* note 70.

232. Or, more technically, this is known as the multiple comparison problem, causing a false discovery rate. More colloquially, one of us once heard this described as "dancing barefoot through a field of data" without any particular goal in mind.

233. One crude adjustment is to reduce proportionately the probability level that is required to show statistical significance. Thus, for instance, assuming the usual level is .05, if one hundred different comparisons are made, each would need to show a probability of chance occurrence less than .0005 to be truly significant, a very demanding test. Less conservative adjustments are also

Potentially more revealing is multiple regression analysis, which measures simultaneously the amount of impact that various possible influences (the independent variables) have on a single effect (the dependent variable). Multiple regression can uncover hidden relationships among multiple factors, or can reveal which of several apparent relationships is the most significant. In attempting to explain the outcomes in a set of cases, for example, several factors may each appear significant by themselves, but when each is held constant, only one or two factors may emerge as the more important predictors of decisions. Sometimes, factors that legal analysts thought were dominant turn out to be red herrings.²³⁴ Alternatively, factors that, standing alone, may not appear significant might emerge as such once the influence of other factors is controlled for statistically.²³⁵

Regression analysis is especially well suited for finding hidden patterns and associations in a complex body of case law in which legal analyses or results depend “extensively on discrete and numerous factors” that occur “in varying combinations in the cases” which courts weigh “without apparent fixed standards.”²³⁶ Conventional analysis, even if thorough and systematic, often cannot untangle the most relevant threads in a complex web of interacting factors, or can do so only speculatively.²³⁷

One-half (67/134) of case content analyses, including 41% (33/80) of those done by lawyers, use multiple regression analysis at least once, and this type of analysis is becoming more common (62% (32/52) of projects published since 2000). Regression analysis was especially prevalent in publications where at least one author was a lawyer with a Ph.D. (67% (20/30)). It is advisable, however, to use regression analysis only for cases that are relatively homogenous, focusing on a single or narrow set of legal issues.²³⁸ Otherwise it may become too difficult to measure and control for all the relevant variables.

Another aspect of statistical analysis worth considering in broad perspective is whether each case (or part of a case)²³⁹ should be given equal weight. The unit of analysis is one of the key differences between conventional doctrinal and case content analysis. Conventional legal analysis focuses on leading cases or on different jurisdictions. Conventional scholars analyze judicial reasoning in key cases and study the patterns of legal rules across

possible, but require greater statistical expertise. Another limitation of bivariate analysis is that, if a number of variables emerge as important each by itself, bivariate statistical testing does not reveal whether each is independently significant in its own right or whether some of these are merely tracking others.

234. See *supra* text accompanying notes 86-88.

235. See *supra* text accompanying notes 117-18.

236. Haar et al., *supra* note 7, at 656.

237. McChesney, *supra* note 76.

238. Hammer & Sage, *supra* note 27.

239. Phelps & Gates, *supra* note 122, at 586, for instance, uses each paragraph of an opinion as the unit of analysis.

jurisdictions. Case content analysts tend to adopt a behavioral model that treats each judicial opinion, judge, or vote as an observation of equal interest. As discussed in Part II, this focus can yield a different set of insights, but it is also limited in its ability to study the law itself, as relevant to lawyers, judges and doctrinal scholars. We must remember, then, that these two methodological approaches define an epistemological divide; they are two distinct ways of talking about and coming to know what judges do in legal opinions.

A strength of the academy is that it can approach legal issues from entirely different perspectives, but it might also strengthen legal scholarship to seek more cross-fertilization between these species of scholarship. Thus, we consider whether content analysis can be modified to become more relevant to understanding case law conventionally, as precedent-establishing legal doctrine. We found only a few examples, but they were instructive. One method analyzes case law using jurisdictions, rather than cases, as the unit of analysis.²⁴⁰ That tends to convert the project from one that studies the behavior of individual judges to one that studies collective behavior resulting in the adoption of particular rules of law.²⁴¹ Another possibility is, in studying a collection of cases, to weight them according to an objective measure of their significance, such as how often they have been cited or followed, or where they stand in a line of precedent.²⁴² A difficulty with this approach is deciding how much weight to assign.²⁴³ Without independent, objective verification of the weight any criteria should carry, the best option is to classify cases into different categories and analyze each separately—such as major versus minor decisions, or leading versus following decisions.

In sum, despite the substantial amount of case coding work that has been done to date, considerable development and innovation lie ahead to find the best uses for this most quintessentially legal means of empirical study.

CONCLUSION

Content analysis does not lead us to the Holy Grail of a true legal science, as early proponents claimed it might—at least not yet. The technique is not yet capable of scientifically analyzing legal principles and precedent at the level that is relevant to practicing lawyers or doctrinal scholars. For that traditional

240. *E.g.*, Morriss, *supra* note 47.

241. *E.g.*, KLEIN, *supra* note 94.

242. *See* WILLIAM M. LANDES AND RICHARD A. POSNER, *ECONOMIC STRUCTURE OF TORT LAW* 315 (1987) (suggesting a systematic analysis of tort law cases, “weighted by the number of citations to each case, such number being a proxy for the importance of the case in shaping the law”); Smith, *supra* note 24 (developing a sophisticated network analysis to rank the importance of decisions); http://prawfsblawg.blogs.com/prawfsblawg/2006/11/the_precydent_s.html (describing a future website that will use Smith’s algorithm as a search engine).

243. Nascent efforts to apply network analysis to the citation patterns among cases may eventually prove fruitful in assigning appropriate weights to different cases. *See* Cross & Smith, *supra* note 72.

enterprise, subjective interpretation of key cases is still the best method, and will likely remain so. Understanding law is no more likely to reduce to a wholly scientific method than interpreting poetry is to yield to a literary science. However, the collective judgment of many legal scholars suggests that content analysis is an especially important adjunct to conventional legal analysis. It enables lawyers to read and analyze decisions more systematically and objectively, when doing so serves their research aims. Any question that a lawyer might ask about what courts say or do can be studied in this fashion. This method renders different kinds of insights that complement those gained through conventional interpretive methods, making the two techniques together more revealing than either alone. Over one hundred and thirty studies and a rapid growth rate amount to a powerful endorsement.

Considering the legal realist underpinnings of most legal scholarship, it should not be a surprise that so many case law researchers have found their way to doing content analysis. This article aims to help others who may wish to follow their lead, realizing that it is not necessary to employ content analysis to its fullest extent and in textbook form. Any legal analyst can profit from basic components such as replicable case selection and coded reading. Some aspects of these techniques are as longstanding and widespread as are West's Key Number System and Shepard's Citations.

Our primary aim here is not prescriptive. Instead, we present what might be called a common law of case coding. We document, describe, and digest how legal researchers have in fact used the technique, and we generalize from these examples, in light of authoritative methods literature, about best practices and applications. As in the actual common law, we anticipate that best practices will evolve as researchers refine techniques and explore new applications. Our hope is that this will now occur with greater awareness of the full range of what other legal scholars and social scientists have already done.

Our survey shows that lawyers' applications of content analysis go beyond those made by political scientists, who study mainly the outcome of cases and their subject matter. Lawyers' applications also go beyond those made by other social scientists, who study the interaction of judge-made law with external social and economic conditions. Legal scholars use content analysis to study the interstices of case law, just as biologists study nature, anthropologists study cultures, or historians study past events. They carefully observe, document, describe, and analyze what they find through an explicit, replicable selection and examination of observable phenomena. Content analysis is the quintessential method for case law empiricism simply because the relevant phenomena for lawyers consist of texts: written opinions.

Content analysis is much more than a better way to read cases, though. It has the power to transform classic interpretive skills into recognizable and transferable social science knowledge. In other words, this method creates a vessel for exporting the analytical insights of legal scholars in a form that the

rest of the social science world will treat seriously. This is also more than just legal scholars adopting scientific methods to study social phenomenon relevant to the law, and more than social scientists studying legal phenomena. Content analysis allows the legal academy to cross-pollinate our understanding of legal principles and institutions with the objective methods and epistemological assumptions of a social scientist. Doing this, legal scholars can lay claim to a specialized ability to apply content analysis to the legally relevant aspects of judicial opinions. To this extent, content analysis forms the basis for a uniquely legal empirical methodology.