December, 2007

Systematic Content Analysis of Judicial Opinions

Ronald F. Wright, Wake Forest University
Mark A. Hall, Wake Forest University

Available at: https://works.bepress.com/ronald_wright/1/
Systematic Content Analysis of Judicial Opinions

Mark A. Hall* & Ronald F. Wright**

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.

Professor Herman Oliphant, in his 1928 inaugural address as President of the
American Association of Law Schools1

Introduction: The Restatement (First) of Coding Cases

Legal scholars are great borrowers of scholarly methods, the mockingbirds of the
academy. We experiment with the tools of historians, economists, sociologists, literary theorists,
moral philosophers, and many others, often to great effect. Yet despite all of these innovative
efforts to study legal doctrines and institutions through different lenses, legal scholars have yet to
identify an empirical methodology that is uniquely their own. Instead, empirical legal methods
are often very 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 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. What many empirical legal scholars
do could be done equally well or better by a social scientist trained in a discipline other than law.

We propose in this Article that one standard social science technique — content analysis
— could form the basis for an empirical methodology that is uniquely legal. On the surface,
content analysis appears simple, even trivial, to some. Using this method, a scholar collects a set
of documents (for instance, judicial opinions on a particular subject), reads the documents

---

* 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.
** Professor of Law, Wake Forest University. Many people gave us useful information and insights, but we are especially grateful to Peter Jacobson, Jason Czarnezki, Jim Brudney, Polk Wagner, and 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).
systematically, records consistent features of each one, and then draws inferences about the use and meaning of those documents. The method comes naturally to legal scholars, because it resembles the classic scholarly routine of reading a collection of cases, finding common threads that link the opinions, and commenting on their significance. But content analysis is much more than a better way to read cases. It could bring social science rigor to our understanding of caselaw, which could create a distinctively legal form of empiricism.

We are not proselytizers of this method. It has certain advantages, along with substantial limitations, compared with conventional legal analysis. What we claim is that, when one chooses to read cases this way, one is engaged 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. Other social scientists can use this or other methods to study questions that relate to the law. Econometricians, political scientists, or sociologists can measure various external drivers and effects of law and legal institutions. But content analysis aims for 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. One early advocate of this method claims that it points us more directly toward the Holy Grail of a century-long quest for a true legal science. Our view is not this grandiose. Instead, we 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 researchers should be able to replicate the results of the research. Put another way, the results of the research matter more than the authority of the researcher.

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, much in the same way that judges developed the common law. Many users of content analysis designed their studies without any reference 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 based to some extent on the technique prescribed in other social sciences, but which differs from the classic approach in some respects. The distinctive concerns of lawyers—as

---

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”).
5 Reed C. Lawlor, Personal Stare Decisis, 41 So. Cal. L. Rev. 73 (1967). For the beginnings of this quest, see, e.g., Roscoe Pound, Mechanical Jurisprudence, 8 Colum. L. Rev. 605 (1908) (calling for a more “exact justice, that is for a justice whose operations within 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 242 (1920).
researchers and as readers—have determined which standard techniques are used and which are neglected.

Like common law judges, the academic users of content analysis have been slow to generalize about their choices. Many have noted 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, perhaps revealing patterns that are more broadly typical. The method also helps a researcher to sort out the interaction of multiple factors that all have some bearing on an outcome in the legal system. Beyond passing observations such as these, 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 use that experience to generalize about the strengths and weaknesses of the technique. What 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? 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. The method also works best when the judicial opinions in a collection hold essentially equal value, such as settings where patterns across cases matter more than a deeply reflective understanding of a single pivotal case.

Content analysis also assumes an equality among readers of judicial opinions. The more traditional legal scholarly enterprise, like the interpretation of literature, relies on the interpreter’s authoritative expertise to select important cases and to draw out noteworthy themes and potential social effects that might flow from the decisions. Content analysis requires the researcher to explain the selection of cases and the search for chosen themes in enough objective detail to allow others to replicate the same steps. This method’s persuasiveness depends on the ability of a community to reproduce the findings rather than the rhetorical power of the author to proclaim them. Content analysis does not displace traditional interpretive legal scholarship, however. 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 this research technique in ways that meet standards of social science rigor yet meet the practical needs of legal researchers. This review takes us through each of the main components of content analysis: selecting cases, coding cases, establishing reliability, and analyzing data. All parts of this article rely heavily on our own content analysis of content analyses. We systematically coded information from all 134 projects we have been able to locate that report on 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, we hope both to describe past practices and to point the way to a better future. The

---

6 In case there is any confusion on this point, this paper is not affiliated with the American Law Institute.
task of a Restatement reporter balances two competing objectives: first, to describe in a systematic way the decisions of courts in a particular area, and second, to highlight promising developments and to encourage more uniform reliance on the most appealing parts of past practice. We aspire in this article to earn a similar dual role, by collecting many 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 ARRIVAL OF CONTENT ANALYSIS IN LEGAL SCHOLARSHIP

Content analysis developed decades ago as a method of systematic reading and analysis of texts. Some of its earliest users worked in the disciplines of sociology and political science, and this method is now used widely in the communications field, as well. Content analysis can be applied to texts of any kind, including legal documents such as trial court records, statutes and regulations. Our sole focus here, however, is the most common use of the technique in legal scholarship: coding judicial opinions.

The first self-conscious use of content analysis to explore the written opinions of judges came from political scientist Fred Kort, in a 1957 article published in a political science journal. Kort collected all the Supreme Court opinions discussing the constitutional right to legal counsel in criminal cases, a total of 28 cases decided between 1932 and 1956. Hoping to explore the

---

7 See also Charles M. Haar, John P. Sawyer, Stephen Cummings, Computer Power and Legal Reasoning: A Case Study of Judicial Decision Prediction in Zoning Amendment Cases, 2 AM. BAR FOUND. RES. J. 651, 742 (1977) (referring to their content analysis as a “computerized restatement” of zoning law).
8 See Richard L. Cupp, Jr., Proximate Cause, Proposed Basic Principles Restatement, and Products Liability, 53 S.C. L. REV. 1085, 1088 (2002) (explaining the role of Restatement reporter); 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, 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.”).
9 See 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); W.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.
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).
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 when trial judges decided whether due process required them to appoint a defense lawyer, and lawyers believed that this jumble of “special 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 in 12 of the 14 remaining cases (86%) in his collection.\textsuperscript{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.\textsuperscript{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 decisionmaking (applying neutral principles of law to the facts of the case) and instead sought to understand the political or personal attitudes that drove judicial behavior.\textsuperscript{15} These early efforts reflected the move toward quantitative methods in political science and the social sciences more generally.\textsuperscript{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.\textsuperscript{17}

A few lawyers and legal scholars began about the same time (in the late 1950s and early 1960s) to develop spontaneously a self-taught method that could 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,\textsuperscript{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,

\textsuperscript{13} Fred Kort, \textit{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.

Kort’s efforts may have been the first self-conscious use of content analysis, but there were earlier precedents. Every lawyer is very 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 http://lawschool.westlaw.com/KNumbers/history.asp?mainpage=16&subpage=4; http://www.lexisnexis.com/shepards/print/; \textit{WEST’S ANALYSIS OF AMERICAN LAW} (rev. ed. 2005); James F. Spriggs & Thomas G. Hansford, \textit{Measuring Legal Change: The Reliability and Validity of Shepard’s Citations}, 53 POL. RES. Q. 327 (2000).


\textsuperscript{16} Krippendorff, \textit{supra} note 3.

\textsuperscript{17} See text at notes \textit{infra} (discussing origins and uses of term “jurimetrics”).

Federico read a more comprehensive set of cases (the 50 most recent examples of Courts of Appeal invalidating a patent) and recorded consistent information about each of them (the stated grounds for invalidity). Others did similar work in the field of medical malpractice, coding and tabulating basic information about almost 2000 published opinions over a span of 160 years.19

During the 1960s and 1970s, the content analysis of judicial opinions expanded in several directions, reaching many 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 other questions of judicial methods also occupied some of the 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 also became larger 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 now 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 responded to a larger number of cases available and the greater facility in accessing them using 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 Westlaw,25

---

20 Werner F. Grunbaum & Albert Newhouse, Quantitative Analysis of Judicial Decisions: Some Problems in Prediction, 3 HOUS. L. REV. 201 (1965); see also Haar et al., supra note 7.
23 See Posner, supra note 11.
as well as software that could perform many of the tedious mathematical calculations involved in statistical analysis.

Legal scholars (including ourselves) came to content analysis in much greater numbers in the 1990s, and its accelerated use has continued into the first few years of the new century.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 8 per year through the middle of 2006.

Looking back from the present day, we can now see that users of content analysis assembled an impressive collection of studies over the years. Many of these projects have been published in obscure places, but many have also been published in the very best law journals.27 Readers of the literature might not appreciate just how widespread the method has become, since most studies only cite to a few examples of this research technique, if they cite 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 were able to locate and coded them for pertinent features, as described in an online Appendix.29 Table 1 summarizes the trends over the years.

26 Similarly, in the communications field, where content analysis is used extensively, one leading journal experienced a 6-fold increase over 24 years in the proportion of its articles using content analysis. KIMBERLY A. NEUENDORF, THE CONTENT ANALYSIS GUIDEBOOK 27 (2002) (attributing big increase in use of content analysis in part to fact that so many good archives of material are now accessible by computer, especially on the internet).


29 Wake Forest Univ. Legal Studies Paper, SSRN Abstract No. 91336, http://papers.ssrn.com/abstract=91336. We attempted to identify all empirical projects whose findings were publicly available 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’ identity. 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 fairly 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 included a similar study that also coded the grounds for finding patent invalidity (obviousness, prior art, etc.).

Our unit of analysis is a case coding project, meaning 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 all other relevant publications, looking both for whether the publication met our inclusion criteria, and whether it cited older publications that might be relevant, which we also reviewed for inclusion and citation to other potentially relevant publications. We also looked forward in time 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 (those that we felt were more likely to be cited by other researchers). These additional instances were also reviewed 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 technique into whatever literatures it led, including those for non-legal disciplines such as 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 reliability or our own coding, we each coded a random selection of 16 of each others’ publications for the factors that could be quantified. 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. 6) Other aspects of the project’s methods or analysis that we thought were notable.

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Projects Coded</td>
<td>10</td>
<td>5</td>
<td>10</td>
<td>57</td>
<td>52</td>
<td>134</td>
</tr>
<tr>
<td>Projects Published per year</td>
<td>0.7</td>
<td>0.5</td>
<td>1.0</td>
<td>5.7</td>
<td>8</td>
<td>2.7</td>
</tr>
<tr>
<td>First Author Based in Law (%)</td>
<td>8 (80)</td>
<td>3 (60)</td>
<td>7 (70)</td>
<td>35 (61)</td>
<td>31 (60)</td>
<td>84 (63)</td>
</tr>
<tr>
<td>Any Author with Ph.D. (%)</td>
<td>5 (50)</td>
<td>4 (80)</td>
<td>7 (70)</td>
<td>41 (72)</td>
<td>27 (52)</td>
<td>84 (63)</td>
</tr>
<tr>
<td>Citation to Social Science</td>
<td>4 (40)</td>
<td>5 (100)</td>
<td>7 (70)</td>
<td>28 (49)</td>
<td>22 (42)</td>
<td>66 (49)</td>
</tr>
<tr>
<td>Methodology (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (maximum) Cases per Project</td>
<td>137</td>
<td>300</td>
<td>190</td>
<td>170</td>
<td>399</td>
<td>252</td>
</tr>
<tr>
<td></td>
<td>(1936)</td>
<td>(5904)</td>
<td>(630)</td>
<td>(8895)</td>
<td>(22,000)</td>
<td></td>
</tr>
</tbody>
</table>
The number of content analysis projects we were able to locate is even more impressive considering that this collection excludes many examples of case-coding empirical work that do not meet our high threshold for true content analysis. We did not include 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. We also did not include the many studies of courts’ citation practices and patterns. Instead, our literature review is based on coding of cases that reaches the substance 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 to study many different subject areas within the law. Our collection includes at least 7 projects (5% of the total) in areas as far-ranging as administrative law, constitutional law, corporate and securities law, criminal law and procedure, contracts, employment discrimination, health law, and torts. A large number of the studies explicitly focus on questions of legal methods, judicial decision-making, and

---

31 See note 71 infra.
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), and four studies address family law (see, e.g., 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); Davis, supra note 25).
The spread of content analysis to study judicial opinions also cuts across disciplinary boundaries. Both social scientists and lawyers use the method to analyze judicial opinions. Among legal researchers, however, about a third have a Ph.D. and so likely have been trained in social science methods. It is demanding to do content analysis thoroughly, and those with scientific training are likely to be both better trained and more highly motivated to take on these painstaking methods. However, it is not necessary to practice this method profitably only at its highest level. Instead, we will show that valuable uses can be made even by those who are largely innumerate.

The payoff for scholars who use content analysis is substantial. Using 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 77 citations per article, with a mean of 30 citations per project. Considering projects from all decades, 87% received at least one citation and 71% generated at least 5 citations. These citation patterns compare favorably to the trends in legal scholarship more generally. According to the ongoing research of Thomas Smith on the citation patterns in legal scholarship, 40% of articles receive no citations at all. 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 to other examples from the legal literature. As Table 1 shows, only about 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 re-invent 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 had discovered something new until we

---


42 See note 27, supra.


44 Our analysis is based on a random sample of 45 projects (one third of the total) covered by the LEXIS-NEXIS database. For each of these, 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 some other article that used content analysis. Thus, we are somewhat less likely to find in our pool articles that received no citations at all.

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 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 “tendency for each new enterprise to invent its own wheel, often in a fairly ad hoc manner”).
learned that we had each done the same thing independently. We see now that many of our colleagues have shared the same sense of finding their own way.

Legal scholars using content analysis often describe their steps at excruciating length that social scientists would view as absurd. It is not uncommon to find articles exceeding 100 pages, describing at length how the author devised various techniques and resolved various quandaries. Others have noted that this has resulted in a “confused area of scholarship, with few clear norms or standards to guide researchers.” This profusion of do-it-yourself research techniques points to a crying need for methodological tenets that can be shared, learned, refined, and cited briefly by legal scholars.

In retrospect, it is not surprising that legal scholars made the transition to content analysis from 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 more than one opinion, because as Karl Llewellyn’s famous introductory text for law students puts it, “no case can have a meaning by itself!” 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 many scholars slipped into doing 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 caselaw.

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

---

48 Hammer & Sage, supra note 27 at 560.
49 KARL LLEWELLYN, THE BRAMBLE BUSH: ON OUR LAW AND ITS STUDY 48-49 (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).
estoppel was overtaking consideration as the basis for enforcing promises, even in the absence of 
detrimental reliance, so he undertook to code all of the relevant cases for two years to find out if 
this was really so (it wasn’t).52

Although the treatise writer and the content analyst have much in common, they each 
offer 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. Researchers who 
perform content analysis, on the other hand, do not audition soloists. Instead, they assemble a 
chorus, listening to the sound that the cases all make together. This distinction between the 
collective and individual insights one might draw 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. Herman Oliphant, a leader 
among the first Legal Realists, is quoted at the outset of this article exhorting law professors to 
ellay 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 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 methods of social science were not as fully developed then.

Perhaps another reason Legal Realists failed to employ empirical methods is that they 
ever really needed to do much of this hard work. They handily won their jurisprudential battle 
against Legal Formalism, so that now it is sometimes said in the legal academy that 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, it is 
completely commonplace when analyzing caselaw to make empirical claims about the actual

52 Hillman, _supra_ note 27.
55 Oliver Wendell Holmes, Jr., _The Path of the Law_, 10 HARV. L. REV. 457 (1897).
56 Stewart Macaulay, _The New Versus the Old Legal Realism: "Things Ain't What They Used to Be,"_ 2005 WIS. L. REV. 365, 374-77; Leiter, _supra_ note 54, at 311-12 (1997); 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 a rich 
array of interdisciplinary social science methods. _See_ Symposium, _Is it Time for a New Legal Realism_, 2005 Wis. L. Rev. 335.
57 Leiter, _supra_ note 54.
thinking and behavior of judges in deciding cases, as partially revealed in their opinions. Using empirical methodology to support these claims means that scholars are only now 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 law respond predictably to various social, political and market conditions—empirical claims that can be tested systematically. Systematic content analysis allows scholars to verify or refute the empirical claims about caselaw 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 the importance of how judges explain their decisions. Although these fundamental differences hugely affect how content analysis might be employed, the method itself 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.

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. Lawyers and legal scholars, however, are often not making empirical claims about a body of caselaw that they intend to be scientifically verifiable. Instead, they mean to analyze particular decisions in a subjective, interpretative, or advocacy mode more akin to how a literary critic might interpret poetry or how a social reformer might critique 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 many assertions about how judges think or act, these are “empirical” only in a casual way. These legal analysts report what they see in key cases and how they interpret these observations, without necessarily making any claims that what they observe is prevalent in the law or typical of other cases. For this kind of scholarship, systematic content analysis would be largely beside the point. Establishing some plausible basis for the minimally empirical claims in

---


59 By positivism, we mean logical positivism, the basic philosophical view of the world underlying modern science, more than legal positivism, which examines the 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 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).

such work is usually done simply by citing relevant sources that readers (or law review students) can verify.

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?” 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.”

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

A. The Components of Content Analysis

There are three distinct components of content analysis: 1) selecting cases; 2) coding cases; and 3) analyzing (often through statistics) the case coding. Each of these components contributes something of value, yet at each of these three stages, some legal scholars have objected to the technique. They question whether content analysis adds anything to traditional interpretive methods, or even 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. The answer to this ground-level challenge to content analysis requires us to compare each of the three components of content analysis with more 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, which are the “leading cases” that best illustrate the historical moment 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 all of the opinions.

---

62 Id. at 300.
A prominent historical example is Karl Llewellyn’s 1960 book *The Common Law Tradition*, which records his observations after systematically reading the “mine run” of thousands of cases randomly selected from courts across the country. He did this to gain a more accurate understanding of what cases courts were hearing and the rhetorical styles they used to decide them. Similarly, Landes’ and Posner’s classic *The Economic Structure of Tort Law* contains multiple examples of systematic case selection, including 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, and every federal appellate products liability case for three years. More recently, conventional non-empirical legal scholars increasingly take advantage of computerized legal databases to specify the universe of cases they research.

We are not necessarily calling on all 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 along the way how they were located, just as a careful historian might do, even if it is not necessary to include the documentation in every publication. This discipline is more likely to uncover all the relevant cases and guards against subliminal biases in selecting only cases that prove the author’s point.

2. Systematic Case Coding

Content analysis leads scholars to read their selected cases more uniformly and with greater focus on their announced research questions. Users of this method must create a coding

---

63 LLEWELLYN, *supra* note 22.
65 E.g., Sidney A. Shapiro & Richard E. Levy, *Judicial Incentives and Indeterminacy in Substantive Review of Administrative Decisions*, 44 DUKE L.J. 1051, 1067 (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 20 of these confirmed that most were instances of authors conducting database searches to verify points made in their articles.
66 See. e.g., Douglas Laycock, *The Death of the Irreparable Injury Rule* 23-24 (1991) (“One 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”).
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 eliminates casual meandering through factors on a case-by-case basis.67

A defined coding scheme focuses attention more methodically on various elements of cases and is a check against looking, either consciously or not, mainly 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 caselaw interpretation.

This degree of rigor in reading cases can be especially helpful when trying to document the absence of some element that is thought to be present in caselaw. A prominent example is Douglas Laycock’s The Death of the Irreparable Injury Rule, which demonstrated based on reading more than 1400 cases that irreparable injury is not an important factor when courts grant or deny equitable relief.68 “Proving a negative” is much harder than simply pointing to what is present in caselaw because the nay-saying researcher needs to demonstrate that he or she has looked systematically for all likely instances of the missing element.

3. Analyzing Cases

At the next stage, information coded from opinions can be tabulated quantitatively, allowing scholars to draw conclusions from the features that they find throughout the cases. At this point, legal researchers enter the realm occupied by social scientists.69 They describe patterns and associations in opinions, with some confidence about the validity and reliability of those observations across the broader universe of cases that they sampled.

We collected over 130 examples of this method in an on-line Appendix.70 They all, using one technique or another, analyze the consistently-recorded factual or legal content of judicial opinions. Many other efforts to analyze some features of judicial decisions do not make the cut because they did not code a sufficient level of detail from the content of opinions to meet our

---

67 Haar et al., supra note 7 at 746.
68 See LAYCOCK, supra note 66.
69 Social scientists also use rigorous qualitative analysis, such as ethnography, but we found only one example based on case content analysis. 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.
70 Note 29 supra.
inclusion criteria.71 We focused our survey on examples of content analysis that required a reader of judicial opinions to bring some legal judgment to bear, such as efforts to describe the content of the parties’ arguments or the judge’s reasoning, or to study the influence of legally-relevant facts.

The precise methods used to quantify cases are subject to a number of criticisms, discussed later in this article.72 Here, we consider a more general challenge to any case coding effort. Some 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.”73

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 (when the constitutional right still depended nominally on the presence of “special circumstances”). Conventional legal scholars, by tuning in to the hints in the most recent in the line of cases, were able to predict this shift correctly.74

The most effective response is to acknowledge the validity of these 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.75 We agree with Karen Jordon,

---


72 See text at notes ___ Part III, infra.


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. 76 These methods also have “considerable power for the discovery of anomalies which may escape the naked eye.” 77 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.” 78 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.

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. 79 Quantitative description can tell us the what of caselaw; 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. 80

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 from one opinion is potentially as relevant as that from any other opinion. 81 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. 82 This is often true in legal analysis because precedent and persuasiveness depend on various qualitative judgments about the reasons given or the source of the decision. 83 “The legal and cultural salience of Roe v. Wade far outruns its statistical significance.” 84

---


79 Levine, supra note 61, at 286.

80 Accord 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”); Lee Epstein, Nancy Staudt, & Peter Wiedenbeck, 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”).

81 Accord, Hammer & Sage, supra note 27; 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 75; Jacobsen et al., supra note 75.

82 See text at notes ___ infra (discussing prospects and difficulties of weighted analysis).

83 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.

84 Goldsmith & Vermeule, supra note 60, at 160.
Taking this limitation into account, scholars have found that it is especially useful to code and count cases in studies that debunk conventional legal wisdom. Sometimes authors even do this to show they were wrong in their own published preconceptions that were based on conventional scholarship. 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. For instance James Henderson and Theodore Eisenberg first documented the sharp shift of judicial attitudes in the 1980s regarding tort liability, showing that products liability litigation was not increasing rapidly as often claimed by critics.

Conventional interpretive analysis is sometimes wrong about what appears in the caselaw or what affects case outcomes. Rather than engaging in a shouting match over whether selected examples amount to a real trend, a far better way to support one’s position is through systematic, objective coding and counting that anyone can replicate using similar methods. The data show X rather than Y or not-X; case closed. Or maybe not, for sometimes the debate continues. 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. This engendered several more case coding projects to refine the relevant data and deepen the debate about the meaning of documented trends. Even though coding and analyzing cases did not end the discussion, the discussion moved to a more observable evidentiary basis. Despite the lack of clear resolution, this series of long articles on promissory estoppel serves as an encouraging example of the legal academy using classic social-scientific methods to critique and correct itself.

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. Some researchers, though, pursue a more structured and ambitious goal. They use sophisticated statistical methods to model or predict in a formal

85 Hall et al., supra note 28; Schuck & Wang, supra note 27.
87 Henderson & Eisenberg, supra note 51.
89 Hillman, supra note 27.
90 See 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).
91 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 text at notes ___Part II B 1, infra.
manner the behavior of judges, the content of opinions, or their relationship with external factors. This general approach at one time was called “jurimetrics.” 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. Other times, the purpose is more scholarly—to test various claims based on legal theory.

These studies of legally relevant factors allow us to draw the most pointed comparison between content analysis and more traditional interpretive methods. Again, we turn to Charles Haar, an early and overlooked visionary of this method, to underscore the potential advantages over traditional legal analysis, in which

[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


93 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, e.g., Allison & Lemley, supra note 86; 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).

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.95

These scholars seek to test precise hypotheses or construct formal models of how complex legal or factual elements affect case outcomes.96 An especially interesting subgenre uses content analysis to find some order and logic in a body of caselaw that, by conventional analysis, appears chaotic or haphazard. As Fred McChesney notes, “the 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.”97 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 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 caselaw.98

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 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.99

95 Haar et al., supra note 7, at 655.
96 For example, this was Fred Kort’s aim in the first conscious application of content analysis to appellate cases. As described above, in text at notes ___, he showed that the outcome of Supreme Court right-to-counsel cases could be predicted reliably from various combinations of factors stated in the opinions.
97 McChesney, supra note 75, at 533.
98 Id. (defective incorporation); Robert B. Thompson, Piercing the Corporate Veil: An Empirical Study, 76 CORNELL L. REV. 1036 (1991) (piercing the corporate veil); Bernard Trujillo, Patterns in a Complex System: An Empirical Study of Valuation in Business Bankruptcy Cases, 53 UCLA L. REV. 357 (2005) (valuation of assets in bankruptcy); Jeffrey A. Segal, Predicting Supreme Court Cases Probabilistically: The Search and Seizure Cases, 1962-1981, 78 AM. POL. SCI. REV. 891 (1984) (search and seizure); Kort, supra note 13 (right to counsel); Kevin T. McGuire, Obscenity, Libertarian Values, and Decision Making in the Supreme Court, 18 AMER. POLITICS QUARTERLY 47 (1990), and Timothy M. Hagle, But Do They Have to See It to Know It? The Supreme Court's Obscenity and Pornography Decisions, 44 WESTERN POLITICAL QUARTERLY 1039-1054 (1991) (obscenity); FRED KORT, CONTENT ANALYSIS OF JUDICIAL OPINIONS AND RULES OF LAW, IN JUDICIAL DECISION-MAKING (Glendon Schubert, ed.) (1963) (involuntary confession); Haar et al., supra note 7 (zoning variances); Ostas, supra note 58 (unconscionability).
99 Juliano & Schwab, supra note 75, at 558-559. See also Krawiec & Zeiler, supra note 27, at 1832 (discussing this concern).
This limitation entails two distinct problems: factual incompleteness, and factual distortion.\textsuperscript{100} Incompleteness results because judges’ factual presentations are meant only to explain as much of the case as is necessary to justify the outcome.\textsuperscript{101} This judicial parsimony can severely distort analysts’ measurement of facts that might be important for research purposes across a range of cases. 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.\textsuperscript{102} Most likely, this is because courts usually do not consider it appropriate to mention race unless they think this might be 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 caselaw, however, need a reasonably accurate indication in each case, not just in those where the judge considered race legally relevant.

The second problem is the possibility that judges distort the facts they report to justify the results they reach. This obviously is a highly contentious charge, but distortion does not have to amount to an outright misrepresentation of facts. It is quite possible that judges do not present opposing facts in as strong a light as they might. Even if this much is true—that the facts used to predict outcomes statistically 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.\textsuperscript{103} Conclusions drawn from statistical analysis of facts presented 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.

One common response to this dilemma is to observe that the incomplete and distorted view of the facts from judicial opinions is “precisely what every lawyer [uses] when reading a case, . . . ; the application of mathematical techniques cannot make the method less valid.”\textsuperscript{104} But

\textsuperscript{100} These also can be thought of as Type II and Type I errors, respectively, or, in testing 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.
\textsuperscript{101} Reed C. Lawlor, Fact Content Analysis of Judicial Opinions, 8 JURIMETRICS J. 107, 109-10 (1966-1968).
\textsuperscript{102} Juliano & Schwab, supra note 75, at 558.
\textsuperscript{103} See Mendelson, supra note 73, 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) (“some 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 ___ 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 78 at 1615.
\textsuperscript{104} Tyree, supra note 77 at 2. Accord Hillman, supra note 27; Theodore Eisenberg & James A. Henderson, Products Liability Cases on Appeal: An Empirical Study, 16 JUSTICE SYS. J. 117 (1993). 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.
statistics cannot make this method of caselaw 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 and 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, both for government work and for 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 don’t say what they really think, they say what they imagine the researcher wants to hear, or they try to be logically consistent across questions even if this distorts the truth. Still, these imperfect data must suffice because observing actual behaviors and gauging true attitudes would be impossible or cost-prohibitive. Similarly, for facts reported by judges, even though they may not “purport to be the real facts,” they are “near enough so that the savings in labor justifies the approximation.”105 This assumption is not heroic. The lawyers and law professors who stake their life’s work on believing (by and large) judges’ renditions of facts are, on the whole, hardly naive idealists.

Second, researchers can specifically 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.106 One interesting technique, used in four studies,107 is to compare 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.108 Such reliance on the adversarial process (here, competing accounts of the facts that appear in

Haar et al., supra note 7, at 743-44. 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 research 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.”  

105 Tyree, supra note 77 at 32. Accord Lawlor, Lawlor, supra note 101 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”).

106 Hillman, for instance, in his prominent study of promissory estoppel, read a sample of cases more closely looking for any indications that courts stated there was reliance where there really wasn’t. 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. Krawiec and Zeiler had students code cases both for whether 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.

107 Richards & Kritzer, supra note 94; Posner, supra note 11; Joseph A. Ignagni, U.S. Supreme Court Decision Making and the Free Exercise Clause, 55 REV. POLITICS 511 (1993); Lawlor, supra note 101 at 115.

108 Lawlor, supra note 101 at 115 is especially thorough. Using student coders, he tested for 30 possible facts in two Supreme Court cases that each had three opinions. Out of these 60 observations that provided 180 possibilities for discrepancies to appear, he found only two.
dissenting opinions) extends into the research arena the lawyer’s classic process-based technique for truth-finding at trial.

Third, the “bias” created by courts’ justifying their decisions may be precisely what a researcher wishes to study. After all, the facts and reasons the judge selects are the substance of the opinion that creates law and binding precedent, so they merit careful study for this very reason. This justification calls, however, 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 predict successes and failures based on the factors present in their clients’ cases, but perhaps is 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 this source of data is simply too valuable to disregard entirely, despite its flaws. Social scientists recognize that judicial opinions “are a tremendous resource,” “probably the best single source of systematic historical data on the law.” Three-quarters of a century ago, Columbia Law Professor Herman Oliphant in his inaugural address as President of the 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.” A half century ago, Karl Llewellyn recognized that the case reporters present an “amazing gathered treasure, … waiting for explorations.” Modern-day legal researchers tend to agree, judging from the large number who have devoted considerable effort to using this source of data.

In sum, the circularity problem is the most serious of the ground-level objections to content analysis. Because of this problem, 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 appeal based on real-world or trial-record views of the facts. But, this limitation is not fundamentally different from similar limitations in many social science studies. Few social scientists use content analysis in any context to draw definitive cause-and-effect conclusions about complex events. Instead, they more often identify apparent associations of interest that merit further study.

109 Barry Friedman, Taking Law Seriously, 4 PERSPECTIVES ON POLITICS 261, 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 98 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 Ignagni, supra note 107; Hillman, supra note 27.

110 See Friedman et al., supra note 71 at 773.

111 See Bliss Cartwright, Conclusion: Disputes and Reported Cases, 9 L. & SOC’Y REV. 369, 383 (1975).

112 Oliphant, supra note 1 at 161. Accord Trujillo, supra note 98 at 363 (stressing the value of the “large amounts of easily detectible data” produced by the bankruptcy courts).

113 LLEWELLYN, supra note 22 at 514.
Moreover, the same limitation applies to traditional interpretive reading of judicial opinions. Both methods face difficulties when their users assume that the facts mentioned in the opinions correspond to the facts on the ground. For both the interpretive method and content analysis, case facts are more appropriately used to identify factors that judges treat as relevant in various bodies of law. We can measure what judges say and do, internal to the caselaw, even if we cannot measure externally what set of raw facts produced those legal results.

B. What Content Analysis Can and Cannot Tell Us

Whether and when to favor content analysis over other methods for studying caselaw is ultimately dictated by one’s epistemological ambitions since it is better suited to answering some research questions than others. We have just seen, for instance, that content analysis is more valid when it attempts to show the significance of factors internal to judicial opinions rather than trying to predict the impact of real-world facts on the litigation. In this section, we explore more generally what research questions are the best targets for users of content analysis. We consider the epistemological fit between this method and projects that aim to learn about 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. Fittingly, the projects that best take advantage of the strengths of content analysis in the hands of legally-trained researchers are those that ask questions familiar to those that traditional legal scholars ask. The different methods produce distinctive insights, but they are well-suited to address similar types of questions.

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, such as political influences, judges’ personal characteristics, or institutional factors (such as court structure). Measuring case results differs fundamentally from measuring the law of a case. Who won and who lost is a much narrower and more objective question, requiring much less legal judgment, than what legal principle a case embodies. It is also a much easier, although not entirely a simple matter, to code just for outcome. Accordingly, this type of study has been done frequently.

The most prominent examples of this genre question the fundamental rule of law in judicial opinions. For over a decade, heated debate has raged among prominent political scientists, law professors, and judges 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. This debate has spawned a vast, often contentious literature—much of it based on empirical studies

114 See Trujillo, supra note 98, and Hammer & Sage, supra note 27, for similar, but somewhat different, categorizations of types of case coding studies.
115 See text at notes infra.
116 For reviews and analysis, see George, supra note 71; Gillman, supra note 73; Sisk & Heise, supra note 2; Emerson H. Tiller & Frank B. Cross, What is Legal Doctrine? 100 NW. U. L. REV. 517 (2006).
using content analysis. A much calmer back-eddy of this type of research studies whether various institutional structures within the judiciary affect case outcomes or judicial reasoning—such as elected versus appointed judges, their length of tenure, the use of specialized courts, crowded dockets, and the like.

These lines of study can be broadened to 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 caselaw as an independent variable, meaning that they ask how caselaw 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. Researchers have neglected the possible effects of judge-made law, including statutory interpretation. With its diverse “laboratory of states,” the U.S. offers boundless opportunities to learn from the “natural experiments” created by the inevitable differences in caselaw among jurisdictions and over time. Doing so, however, requires accurate and replicable coding of the caselaw.

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. Studying Judicial Reasoning


120 An important resource for such studies is Morriss, supra note 47.
Cases are objects of study for legal scholars 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.\(^{121}\) It is doubtful that content analysis standing alone will be able to replace this core mode of caselaw analysis.

We have seen, though, that content analysis can augment conventional analysis by identifying previously-unnoticed patterns that warrant deeper study, or sometimes correcting misimpressions based on more 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 many 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 caselaw. Content analysis tolerates large elements of imperfection. As described more below, its methods require only acceptable levels of replicability and statistical association.\(^{122}\) The epistemology implicit in much social science is that ultimate reality is only subject 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.\(^ {123}\)

These two points of contrast—the depth and the objectivity of understanding—are related. Social scientists observe that there is often a trade-off between the usefulness of findings and a method’s objectivity, in that more perfectly replicable methods tend to produce narrower, more superficial, and therefore less meaningful findings.\(^ {124}\) 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. Using systematic coding can increase internal validity by removing elements of researcher bias and improving thoroughness and accuracy. However, to the extent that content analysis cannot reach important aspects of legal interpretation that are impossible to code objectively (such as nuance related to infrequent or highly complex factual and procedural patterns), plain content analysis loses relevance, i.e., external validity. When uniform coding cannot capture important details about idiosyncratic decisions, content analysis alone is not capable of measuring what lawyers or scholars would consider to be a full and accurate statement of the law.

\(^{121}\) See generally BENJAMIN N. CARDozo, THE NATURE OF THE JUDICIAL PROCESS (1921).

\(^{122}\) Text and notes ___See Part III, infra. See also KLEIn, supra note 117; Juliano & Schwab, supra note 75 at 557.

\(^{123}\) See generally RONALD Dworkin, LAW’S EMPIRE (1986).

\(^{124}\) See Krippendorff, supra note 3 at 212-13 (2nd ed. 2004). Legal scholars have noted the same point, e.g., Fuller, supra note 78 at 1626-28 (contrasting narrow prediction of case outcomes with deeper understanding of judges’ human behavior and motivation).
b. Normative Legal Analysis

Accordingly, very few opinion-coding projects explicitly aim to evaluate the legal correctness of judicial opinions.\footnote{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. Reviewing these studies, see Neil Vidmar, *Are Juries Competent to Decide Liability in Tort Cases Involving Scientific/Medical Issues? Some Data from Medical Malpractice*, 43 EMORY L.J. 884 (1994); Philip G. Peters, *The Role of the Jury in Modern Malpractice Law*, 87 IOWA L. REV. 911 (2002). However, these studies often struggle with obtaining acceptable reliability of expert professional judgment. See, e.g., Karen L. Posner, Robert A. Caplan, & Frederick W. Cheney, *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). See, e.g., 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, 1900-1911*, 41 CATH. U. L. REV. 325 (1992); Willy E. Rice, *Federal Courts and the Regulation of the Insurance Industry: An Empirical and Historical Analysis of Courts' Inefffectual Attempts to Harmonize Federal Antitrust, Arbitration, and Insolvency statutes with the McCarran-Ferguson Act--1941-1993*, 43 CATH. U. L. REV. 399 (1993); Rice, *supra* note 58 (expressing “anger” and “outrage” at the “astonishing” and “startling” 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.”).} Almost all projects we reviewed are primarily descriptive or analytical rather than normative. Certainly, many content analysts draw normative implications from what they observe,\footnote{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.} but their coding of cases aims only to document what judges do rather than to judge 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 its judges are legally wrong, in an empirical sense? What they say is the law.\footnote{LLEWELLYN, *supra* note 22.}

Among the few exceptions, 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).\footnote{Spriggs & Hansford, *supra* note 13.} Similarly, two political scientists verified the accuracy and reliability of Shepard’s citation service using a random sample of 602 opinions.\footnote{John Gruhl, *The Supreme Court's Impact on the Law of Libel: Compliance by Lower Federal Courts*, 33 WEST. POL. Q. 502 (1980).} Only one project squarely evaluates the legal correctness of judicial opinions, by coding for whether lower courts correctly applied or interpreted the Supreme Court’s libel law precedent, *New York Times v. Sullivan*.\footnote{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. KLEIN, *supra* note 117; Johnson, *supra* note 117. The first author notes the large element of subjectivity in this coding but does not test its reliability. KLEIN, *supra* note 117. The second author reports that there were too few instances of unambiguous noncompliance with precedent to include them in the analysis. Johnson, *supra* note 117 at 327.} 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.\footnote{126 See, e.g., 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, 1900-1911*, 41 CATH. U. L. REV. 325 (1992); Willy E. Rice, *Federal Courts and the Regulation of the Insurance Industry: An Empirical and Historical Analysis of Courts' Inefffectual Attempts to Harmonize Federal Antitrust, Arbitration, and Insolvency statutes with the McCarran-Ferguson Act--1941-1993*, 43 CATH. U. L. REV. 399 (1993); Rice, *supra* note 58 (expressing “anger” and “outrage” at the “astonishing” and “startling” 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.”).}
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—claims that obviously are fraught with jurisprudential difficulties.\textsuperscript{132} Considering these limitations of content analysis, normative evaluation of legal doctrine can be done convincingly only when a project includes a traditional interpretive component.

3. Exploring the Landscape of Caselaw

We are presented, then, with a Goldilocks dilemma – one where the best solution is not too much, or too little, but just right. 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 are correct to 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.”\textsuperscript{133}

Accordingly, most projects we reviewed attempt to take full advantage of the substance of judicial opinions, while avoiding normative legal analysis, by using their factual, rhetorical, and legal details to describe or explore a body of caselaw, 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, much like 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 fairly definitive conclusions about cause-effect relationships that have theoretical significance. Descriptive or exploratory studies are more akin to mapping than to testing.

Descriptive/exploratory studies can be subdivided further into two broad classifications: a) those that examine the background of legal doctrines, case subject matter, or case outcomes, versus b) 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 caselaw and the factors that appear important to case outcomes. A dozen

\textsuperscript{132} It is interesting to contrast clinical medicine, where researchers have developed replicable, scientific methods for evaluating the “appropriateness” of treatment decisions, using panels of expert physicians. \textit{See infra} notes 205-209. 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.

\textsuperscript{133} Friedman, \textit{supra} note 109 at 266. \textit{Accord} Epstein et al., \textit{supra} note 80 at 320-23 (explaining why coding projects should focus on more than mere case outcomes).
explore a variety of civil rights topics, another dozen explore various business law topics, and a half dozen others address issues in health care law. Many focus on who wins and loses and why, but others simply want to understand what is happening in the courts. Others ask what impact a new precedent, statute, or legal doctrine is having. Most prominent among these are a passel of articles exploring the impact of the Supreme Court’s Chevron decision regarding judicial deference to administrative agency decisions.

The primary criticism of these descriptive or exploratory studies is that they sometimes draw conclusions about features of the legal landscape that cannot be observed fully from judicial opinions. As discussed more below, win/loss records from published opinions do not

---


136 Hall et al., supra note 28; Jacobsen et al., supra note 75; Hammer & Sage, supra note 27; 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.

137 For instance: how courts define securities (Gabaldon, supra note 135), and what kinds of cases have state supreme courts decided over the course of a century (Kagan et al., supra note 24).


139 Revesz, supra note 27; Kerr, supra note 94; Cross & Tiller, supra note 27; Merrill, supra note 27; Schuck & Elliot, supra note 25; Herbert M. Kritzer, Mark J. Richards, & Joseph L. Smith, Deciding the Supreme Court’s Administrative Law Cases: Does Chevron Matter? Presentation to American Political Science Association meeting (2002); Christopher H. Schroeder & Robert L. Glicksman, Chevron, State Farm, and EPA in the Courts of Appeals During the 1990s, 31 ENVIRONMENTAL L. REP. 10371 (2001).

140 Text at notes ___ infra.
necessarily tell us about legal disputes that were never filed in court, or 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 because they reveal 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.” \(^\text{141}\) Therefore, as Bernard Trujillo puts it, published opinions are especially probative on questions about the spread of ideas within the legal system:

> [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.\(^\text{142}\)

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;\(^\text{143}\) the argumentative, interpretive, or expressive techniques judges use in different circumstances;\(^\text{144}\) and the various meanings for important social conceptions (such as family or equality) that their written opinions tend to reflect.\(^\text{145}\) This category includes some of the most inventive uses of case content analysis, such as Laura Little’s study of different linguistic devices used to obscure meaning in Supreme Court opinions.\(^\text{146}\)

The largest grouping in this general category are the half dozen authors who analyze courts’ reliance on social policy or social science, in contrast with more technical or formalistic doctrinal reasoning.\(^\text{147}\) Naturally, some caution is warranted in concluding that a mention of some source in an opinion indicates the actual importance judges place on this type of evidence.

\(^{141}\) See Friedman, et al., supra note 71 at 773. Accord Friedman, supra note 109 at 266.
\(^{142}\) Trujillo, supra note 98 at 364-65.
\(^{144}\) Schneider, supra note 117; Epstein et al., supra note 80; Smith & Tiller, supra note 117; Schacter, supra note 27; Sisk et al., supra note 27; Charles A. Johnson, *Content-Analytic Techniques and Judicial Research*, 15 AM. POLITICS Q. 169 (1987); 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); Daynard, supra note 21; Wagner & Petherbridge, supra note 27; Little, supra note 50; Henderson & Eisenberg, supra note 51; Czarnecki & Ford, supra note 118; Rebell & Block, supra note 11.
\(^{146}\) Little, supra note 50.
\(^{147}\) Davis, supra note 25; Brudney & Ditslear, supra note 143; Fradella, supra note 143; Falk, supra note 143; Phelps & Gates, supra note 144; Daynard, supra note 21; Rebell & Block, supra note 11.
and argument. Still, with appropriate caveats on the claims being made, systematic study of how judges reason in their written decisions is perhaps the most compelling application of case content analysis because it best fits the method with the type of question that researchers are asking.

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 caselaw that 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 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 caselaw.

The major limitation of content analysis (a limit that applies equally to traditional interpretive methods) is that facts and reasons given in opinions cannot be treated as accurate and complete. 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. 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 nuanced understanding gained through the interpretive method is better suited for evaluation of legal principles. On the other hand, content analysis 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. Then, 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

148 But, then again, neither are conventional legal methods. In fact, statistically 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 Supreme Court Decision-Making, 104 COLUM. L. REV. 1150 (2004).
So far, we have considered whether and when to use content analysis at all. In this part, we assume this choice has been made, so we next 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 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.\textsuperscript{149}

Our guidelines are based on a review of the universe of case-coding projects (summarized in Table 2), recommendations from leading content analysis methodologists,\textsuperscript{150} and our own experience in coding and counting cases.\textsuperscript{151} We revisit the same three stages of content analysis discussed in Part II A — selecting, coding, and analyzing cases — but now the focus is on how to rather than whether to undertake each task.

\begin{table}[h!]
\centering
\begin{tabular}{|l|c|}
\hline
Percent of Case Coding Projects With: & \\
\hline
 Replicable Case Selection & 87 \\
 Universal Sample & 85 \\
 Student Coding & 22 \\
 Specific Coding Instructions & 60 \\
 Reliability Testing & 14 \\
 Only Descriptive Statistics & 34 \\
\hline
\end{tabular}
\caption{Table 2}
\end{table}

A. Selecting Cases

1. Sampling Frame and Biases

The first decision in any empirical project is 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.

\textsuperscript{149} See Sisk & Heise, 99 Nw. U. L. Rev. 743, supra note 2 at 792.
\textsuperscript{151} Wright & Huck, supra note 28; Hall et al., supra note 28.
The selection method is comparatively easy to handle for most content analyses of judicial opinions. The most common approach we observed was to study 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 stays small enough to manage. In our review, 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 caselaw 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: 1) true random sampling (best done by computer-generated list of random numbers); 2) systematic sampling, such as every fifth case; 3) quota sampling, such as all cases up to 200, per jurisdiction per year; or 4) purposive sampling, such as cases that are cited by leading treatises and casebooks or by other cases. We observed examples of each of these methods.

For case analysis, the more troubling question is the relevant 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. A more diverse selection of cases may be 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 to focus on appellate decisions when analyzing factors that predict success or failure. Others discuss choices between intermediate and highest appeals courts, between state and federal courts, and between published and unpublished opinions.

152 See, e.g., Allison & Lemley, supra note 86 (noting that researchers avoided issues of sampling bias by coding the universe of all 299 relevant cases).
153 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); Schneider, supra note 117; Ross, supra note 134; Benesh & Martinek, supra note 94; Johnson, supra note 144; Falk, supra note 143; Kagan et al., supra note 24; GLORIA K. KOENIG, PATENT INVALIDITY: A STATISTICAL AND SUBSTANTIVE ANALYSIS (rev. ed. 1980); Spriggs & Hansford, supra note 13; Posner, supra note 11; Dixon & Gill, supra note 138; George & Sheehan et al., supra note 24; Brace, supra note 24; Scott, supra note 27
155 E.g., Hammer & Sage, supra note 27; Allison & Lemley, supra note 86; Songer & Haire, supra note 94.
156 Franklin, supra note 24.
157 Songer & Haire, supra note 94; George & Sheehan, supra note 24.
158 Kagan et al., supra note 24; Brace & Butler, supra note 24.
159 Sisk et al., supra note 47.
The choice of sample frame depends critically on the study’s central questions and purposes. Study questions can be narrowed to fit the sample frame that is available, or a theoretical sample frame can be imagined that is unrealistically broad, but that fits a more interesting or important set of questions the analyst wishes to pursue. Political scientists, for instance, often study political and institutional influences on judicial decisionmaking by focusing on a particularly controversial or value-laden set of Supreme Court decisions.159 Observing factors that influence free speech160 or death penalty161 cases, for instance, provides insights into judicial behavior more generally.

Too often, content analysts (including ourselves) have chosen sample frames without conscious reflection. One of us, for instance, chose to explore how courts determine the effectiveness of medical treatment in health insurance disputes, by studying all published judicial opinions resolving such disputes.162 That universe of observations might be relevant for a narrower question, such as how appellate courts reason 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.

Whenever the actual cases selected do not fully match the sampling frame that theoretically applies to the questions posed or studied, an issue of sampling bias potentially exists. The findings from cases studied may not accurately reflect 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.163 Likewise for studies that sample from certain 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.164 Most frequent and extensive 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: only

159 Legal scholars have done this as well, looking at judicial review of agency decisions. See supra notes 94, 139.
160 Richards & Kritzer, supra note 94; Songer & Haire, supra note 94; Gruhl, supra note 130.
161 George & Epstein, supra note 117.
162 Hall et al., supra note 28.
163 See, e.g., Allison & Lemley, supra note 86 (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 example, Mendelson (supra note 74 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 (see text at notes supra) would have ill-served a lawyer who later might have advised a convicted client against appealing based on Kort’s analysis. 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. Wainright (which overtly changed the law). See LEWIS, supra note 12 (discussing this silent shift in the Court’s right-to-counsel jurisprudence).
164 E.g., Revesz, supra note 27; Hammer & Sage, supra note 27; Allison & Lemley, supra note 86; Colker, supra note 134; Schuck & Wang, supra note 27; Sisk et al., supra note 47. Some, however, completely ignore the issue (e.g., Absher & Baker, supra note 93), or misunderstand it, when it is fundamental to the questions they are analyzing. Professor Rice in several of his studies (e.g., Rice, supra note 126; Rice, supra note 134), claims to test statistically for whether there is a selection bias by analyzing whether his findings differ significantly between lower court and appealed cases. However, he is not able to test for whether reported or litigated cases differ from those that are settled or not reported, since, by definition, those cases cannot be observed.
some human interactions 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 of these junctures, there
are a variety of factors that potentially distort what one stage can reveal about the other.  
These biases can fundamentally threaten the validity or generalizeability 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 should not cause big worries.  
No concern arises if the researcher defines the research questions in terms that match the population of cases
actually sampled. For instance, if it is precedential law that one wants to study, rather than
simply the generalized behavior or attitudes of judges, then unpublished opinions are irrelevant
and so excluding them requires no justification. In other situations, where excluded cases are
theoretically relevant, the exclusion can easily be justified if the likely direction of bias or
distortion or considered. 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 produced a
false observation of no effect.

In other situations, likely differences between studied cases and omitted cases are
sufficiently trivial that the omission should create no more a concern than other limitations
obvious and inherent in the sample frame itself, such as one date range rather than another.  
All empirical studies are imperfect—observational (non-experimental) social science studies
especially. 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 only briefly.

---

165 For other examples surveying the literature and documenting various of these points, see Friedman, supra note
109 at 270-72; Kevin M. Clermont & Theodore Eisenberg, Litigation Realities, 88 CORNELL L. REV. 119 (2002);
REV. 501, 502-04 (1989); 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); 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.
166 Accord, Trujillo, supra note 98 at 365-66.
167 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.
168 Juliano & Schwab, supra note 75 at 555, offers 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.
In sum, as long as researchers do not extend their claimed findings beyond the universe of what they have actually studied, they should not agonize over the fact that their chosen sample frame is not capable of answering all conceivably relevant questions conclusively—especially when they have selected all cases within their defined frame. Only when researchers make strong claims about the generalizability of their findings beyond the sampling frame, to a larger population that we have good reason to believe might be different than the one they studied, do they need to go to any length in their publications to consider the potential selection bias and validity problems.\textsuperscript{169}

2. Selection Techniques and Replicability

An essential attribute of scientific objectivity is the ability, at least in theory, to test a research project by reproducing its findings using the same methods.\textsuperscript{170} Replicability is the overriding reason for using systematic content analysis. This is what confers scientific status on the findings and conclusions.

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.\textsuperscript{171} Full replication is precluded, for the most part, only by the subjective judgments used to narrow the initial case selection to those that are directly relevant to the research question. Most case coding projects in the past 15 years have taken advantage of the ability to select cases using structured computer searches in the Westlaw or LEXIS databases. It is good technique to specify the exact search terms and strategies, so these can be replicated by other researchers. Most studies we reviewed did so, but many studies still fell short of a fully replicable case selection.\textsuperscript{172} This is because computerized searches are rarely refined enough to narrow the sample to relevant cases. Cases that mention a topic of interest often do so only in passing. Those that decide an issue sometimes do so on technical or procedural grounds that are not relevant to the study.

When the principal selection technique produces many more cases than merit the effort of content coding, it is necessary to screen initial selections. Most legal researchers do so using criteria of relevance that cannot be fully replicated. Another option, however, is to refine the mechanical search strategy. Among the useful strategies we observed in our literature review are searching case digests or headnotes rather than the full case itself, or searching a sample of cases

\textsuperscript{169} Thus, for instance, Czarnezki & Ford (\textit{supra} note 118 at n.91), worry excessively (saying this is a problem “of great concern”) that running the exact same search in Lexis on different days turned up slightly different case counts with no apparent reason.

\textsuperscript{170} See Epstein & King, \textit{supra} note 2 at 38, 83 (stressing the scientific importance of replicability).

\textsuperscript{171} We coded at least partial replicability for 87% of projects (117/134), but our intercoder agreement for this measure was not reliable (see on-line Appendix, \textit{supra} note 29), so we present this observation only as a qualitative assessment.

\textsuperscript{172} In this context, full replication is not achieved simply by listing all case citations or providing the researcher’s database. \textit{Cf.} Morriss, \textit{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.
selected initially because they cited particular statutes, or because they 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.\textsuperscript{173}

We found no study that formally tested the replicability (or statistical reliability) of the judgmental criteria used to select cases. This was not even done by researchers (such as ourselves) who used state-of-the-art methods to test the reliability of their case coding. Verifying the replicability of case selection would make a study more scientifically rigorous, eliminating the possibility that a researcher subconsciously chose cases according to whether they appear to support the researcher’s preliminary hunches. Either formal reliability testing, or case selection by someone who is otherwise uninvolved in the study, are ways to guard against this potential bias. Failing to do this, however, is often not a major flaw because, in our judgment, the selection criteria in many studies are reasonably objective and so would likely produce a similar selection if employed by other researchers.

B. Coding Cases

Once cases are selected, experts in content analysis outline four basic steps that should be followed in coding any material\textsuperscript{174}:

- Based on the questions that seem most germane to the study, create a tentative set of coding categories a priori. After thorough evaluation, including feedback from colleagues, study team members or expert consultants, refine these categories.
- 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.
- Add, delete, or revise coding categories based on this pilot experience, and repeat reliability testing and coder training as required.
- 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

\textsuperscript{173} Discussing this point, see Phelps & Gates, \textit{supra} note 144.

\textsuperscript{174} Krippendorf, \textit{supra} note 3, at 129-30; Neuendorf, \textit{supra} note 26 at 50-51. A summary flowchart can be found at: \url{http://academic.csuohio.edu/kneuendorf/content/resources/flowc.htm}. 

38
are present in the case in some fashion. 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.

Regarding the bases for decision, 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. A common focus of coding is also the court’s style of analysis or approach to statutory or constitutional interpretation, categorized in various ways.

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. For instance, Charles Johnson demonstrates the ability of law students 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.

Coding experts advise researchers to create more coding categories, and to make coding more fine-grained, than the categories they may ultimately use. Even though this produces more information than the project will eventually require, the advantage is allowing 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, while keeping mutually exclusive categories—in other words, to capture all the relevant information, but to avoid having categories that duplicate or overlap each other. 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.

A good example of exhaustive and mutually exclusive coding is categorizing case outcomes. It is usually not a simple matter to define what counts as a win or loss across a range of cases. Appellate cases arise in a variety of procedural postures, they usually 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 United States Court of

175 E.g., Kritzer & Richards, supra note 94; Posner, supra note 11; Ignagni, supra note 107; Absher & Baker, supra note 93; Johnston, supra note 27; Ostas, supra note 58.
176 E.g., Schuster, supra note 134; Brudney & Ditslear, supra note 75; Jacobson et al., supra note 75; Henderson & Eisenberg, supra note 51.
177 E.g., Schneider, supra note 117; Schacter, supra note 27; Brudney & Ditslear, supra note 75; Phelps & Gates, supra note 144; Daynard, supra note 21; Czarnecki & Ford, supra note 118; Rebell & Block, supra note 11.
178 See NEUENDORF, supra note 26 at 23; Little, supra note 50.
179 Johnson, supra note 144. Some researchers have clearly stretched this too far, however. For instance, Falk, supra note 143, used a 7-point Likert scale to score opinions for their degree of reliance on social science evidence, and Groscup et al., supra note 138, 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. Note _ infra.
Appeals Database, for instance, defines all possible case outcomes using nine categories. This illustrates that it is a better practice to be overinclusive at the coding stage, waiting until the analysis stage to collapse the various categories into discrete win/loss columns.

When categories are finalized, it is essential to good coding practice to record their description and instructions for their application. Obviously, this is necessary if coding is done by someone other than the researcher, such as student assistants. Even if authors do their own coding, the scientific standard of replicability requires a written record of how categories were defined and applied. In our literature review, there was evidence in publications from 60% (80/134) of projects that a codebook was created, whether or not one is actually available now. 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.

Experienced coders advise that errors will be reduced if coding forms are designed to minimize writing. 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. Also, a balance should be struck between a form that is so spare it offers almost no instructional information for coders, forcing them to refer frequently to the detailed coding manual, versus a coding form that is overlong because each form contains a full set of instructions. 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. 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 used student coders primarily; in the rest, the authors appeared to do their own coding. In theory, the most scientifically rigorous method is for researchers to train

---

181 Songer & Sheehan, supra note 24. Discussing this coding variable in particular, see Lee Epstein & Andres Martin, Coding Variables, in ENCYCLOPEDIA OF SOCIAL MEASUREMENT (Kimberly Kempf-Leonard, ed.) (2005).
182 Sometimes the publication itself was sufficiently detailed that its text might be used as a codebook, or the codebook and coding form was included as an appendix.
183 Available at http://www.law.wfu.edu/x5649.xml. Researchers who wish to contribute their materials should contact the current editor in chief, found at http://www.law.wfu.edu/lawreview.xml.
184 See Krippendorf, supra note 3 at 148.
185 Examples of well-designed code books and coding forms can be found at Hammer & Sage, supra note 27; Colker, supra note 134; Lawlor, supra note 101; and the research web pages linked to http://epstein.law.northwestern.edu/cv.php.
186 Lawlor, supra note 101 at 129.
187 Sometimes, this was not entirely clear, but unless there was clear evidence of coding 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
others to do the coding and for coders to work completely independently once they are trained.188 (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.189 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.190 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.191 Still, this researcher found that coding reliability improved the more that students were trained.192 Researchers who use students should describe how training was done in sufficient detail that others could replicate all of these steps.193

Other considerations might counsel making the opposite choice, however. Training coders to achieve accurate and reliable results can be a difficult and time-consuming undertaking, one that may require considerably more resources and effort than researchers simply doing their own coding, especially in smaller projects.194 A relevant selection of cases is often sufficiently small that a single reader can handle the coding alone. In our review, 29 projects (22%) coded fewer than 100 opinions and another 28 (for a total of 42%) coded fewer than 200.

Several researchers we read or spoke to emphasized the surprising number of mistakes student coders can make, even on seemingly simple and objective criteria such as dates.195 We

---

188 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 144 at 175.

189 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.

190 According to Krippendorff, supra note 3 at 218, “[s]elf-applied recording instructions are notoriously unreliable.”

191 Lawlor, supra note 101 at 119.

192 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 Krippendorff, supra note 3 at 129.

193 Good examples are Hammer & Sage, supra note 27; Johnson, supra note 144 at 181-82; and Czarnezki & Ford, supra note 118.

194 Several researchers emphasize how much more difficult this undertaking proved to be than they first anticipated. E.g., Bliss Cartwright, supra note 111, 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.”); Kagan et al., supra note 24, at 128 (coding was “an immense, almost intractable task”); Haar et al., supra note 7, at 656; Lawlor, supra note 101, at 116-22 (describing extensive efforts to train coders, which improved their performance but never achieved perfection); Phelps & Gates, supra note 144 (describing extensive pre-testing, which took 3 months, and coding that took 18 months).

195 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 150-page article developing and testing 18 general principles about how to date changes in the law announced through judicial decisions.
observed several instances of researchers who first began, but then abandoned or altered, coding by students because of the difficulties and frustrations. 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 “dumbed down” to be meaningful. It may be that law student coders do not have 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 still to conduct reliability tests, by recruiting a colleague with similar expertise to double code at least a subset of cases.

The dilemma over who should code goes to the core of the epistemological differences between traditional and empirical methods of legal scholarship. Traditionally, cases are understood best when they are interpreted by the scholars with the most expertise. When each case must receive an individualized reading to develop the scholarly thesis, obviously the scholar should do the reading and interpretation for herself, with students helping only to locate the relevant cases. From a social science perspective, however, doing scholarship this way is the height of unmitigated subjectivism—the opposite of good scholarship. Scientifically, scholarship 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 will be judged incompetent by both camps.

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.” If coding categories are so objective and straightforward that it is obvious they can be applied consistently, 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

---

196 E.g., Colker, supra note 134; Winkler, supra note 134.
197 See Krippendorf, supra note 3 at 212-13 (discussing the trade-off between reliability and validity).
198 In our literature review, this was done in only 6 of the 90 projects where authors did some or all of their own coding. Jacobson et al., supra note 75; Phelps & Gates, supra note 144; James F. Spriggs & Paul J. Wahlbeck, Amicus Curiae and the Role of Information at the Supreme Court, 50 Pol. Res. Q. 365 (1997); Songer & Haire, supra note 94; Spaeth, supra note 24.
199 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 caselaw, NIH scientific 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.
200 See 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. 587-604 (2002).
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. This claim is undermined if coding reflects primarily the subjective, idiosyncratic interpretation of the particular individuals who read the cases, or if coding has large elements of error or arbitrariness.

Nevertheless, we do not mean to say that coding projects cannot be published without testing reliability formally and statistically. Even without any reliability testing, it is perfectly possible that a coding scheme in fact is reliable; as long as the coding is described in sufficient detail for others to replicate, a researcher can choose to let others challenge or verify reliability. 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 formal, quantitative assessment of coding reliability. Legal empiricists should do better, especially considering the inherent subjectivity of reading cases.

It is good technique for content analysts to assure themselves of coding reliability in some fashion. The best method is to conduct formal reliability tests during (at least) two stages in the process: initially, while piloting the draft coding process, and later, once coding categories and instructions are completed. Formal testing calls for at least two coders independently to code a sample of cases and to compare their results statistically. The most common statistic is simple percent of agreement. However, a simple percentage does not account for the level of agreement that would be expected purely by chance. Because chance agreement varies according to the type of coding scheme (that is, a variable with two possible answers will naturally produce more agreement than a variable with eight possible answers), the

---

201 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 Danes & Seaton, supra note 94 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.

202 Surprisingly, these percentages were not dramatically different within the subset of 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 200; Krippendorf, supra note 3 at 142.

203 Model discussions of reliability testing can be found, e.g., in Juliano & Schwab, supra note 75, at 598-602; Gerard F. Anderson, Mark A. Hall, & Teresa R. Smith, When courts review medical appropriateness, 36 MEDICAL CARE 1295, 1297 (1998); Richards & Kritzer, supra note 94; Lawlor, supra note 101.

204 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 9 students averaged 82% after only a brief orientation and training session, but it ranged from 73% to 90% across students. After extensive training, reliability improved to an average of 95%, ranging much more tightly from 94% to 97% across students. Lawlor, supra note 101 at 116-22.

205 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 30, whichever is less. Lombard et al., supra note 200; Neuendorf, supra note 26 at 158-59. For our coding of content analysis databases, we double-coded 32 of the 134 databases (24%).

206 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) half the time. Neuendorf, supra note 26 at 150.
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” (after its inventor) or simply the kappa statistic.\textsuperscript{207} Ranging from 0 to 1, kappa indicates the proportion of observed agreement that exceeds what would be expected by chance alone, with 0 indicating agreement entirely by chance and 1 indicating perfect agreement.\textsuperscript{208}

If statistics such as these are used, they must be employed 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 and so aggregate statistics can mask serious problems with key variables.\textsuperscript{209} Also, when the response pattern for a variable is highly skewed (say, one of several available responses occurs much more frequently than the others), this should be noted or taken into account in the statistical measure used. Otherwise, the nominal level of agreement can be deceptive. Haar and colleagues, for instance, coded for presence or absence of 167 different factors in each case.\textsuperscript{210} 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, however, is whether coders agree when they indicate a factor is present.\textsuperscript{211}

For content analysts who do not have access to statistical testing, another alternative is to report the simple percent 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.\textsuperscript{212} Coders may be surprised how poor the nominal percent of agreement turns out to be. In our review of content analysis publications, for instance, we coded for

\textsuperscript{207} An older statistic is Scott’s pi, and a newer one is Krippendorf’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, \textit{Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research Projects}, (available at http://www.temple.edu/mmc/reliability).

\textsuperscript{208} See NEUENDORF, supra note 26 at 150; Lombard et al., supra note 200. A negative kappa statistic indicates less agreement than expected by chance.

\textsuperscript{209} Lawlor, supra note 101 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 on-line Appendix, supra note 29.

\textsuperscript{210} Haar et al., supra note 7.

\textsuperscript{211} 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 144 at 183; Spaeth, supra note 24 at xviii.

\textsuperscript{212} For an example of this approach used in the review of medical records, see A. Russell Localio, Susan L. Weaver, J. Richard Landis, et al., \textit{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”).

\begin{flushright}
44
\end{flushright}
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. Most of our 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, problems that can be corrected if the error appears after the pilot phase rather than after the completed coding. After final coding, compulsive legal researchers might try to get to the bottom of remaining disagreements and resolve all discrepancies, both in the reliability testing sample and across the entire selection. 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.

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 best practice would be to retest the refined rules for reliability, producing a never-ending cycle in search of elusive perfection. 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.

Although there is broad agreement on the desirability of testing for reliability, and some agreement on the methods for doing so, there is not firm agreement on what level of reliability is acceptable. The goal is aspirational — to achieve high levels of agreement — rather than merely to rise somewhat above purely random agreement. One expert says that a reliability coefficient of .8 (that is, data agree 80% more than mere chance agreement) is good, with indices from .67-.8 being sufficient for “tentative conclusions.” Others claim that this is too demanding, especially for coding categories that produce more skewed responses, since even small levels of disagreement can cause the statistical index to drop rapidly. Therefore, other authors provide

Similar to Lawlor, supra note 101 at 124, our discrepancies in this and other categories were due to a mix of simple errors, incomplete coding instructions, and inherently ambiguous or judgmental coding categories.

Lawlor, supra note 101, 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.” 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.

Still, when disagreements are identified by reliability testing, 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 at al., supra note 200.

See KRIPPENDORF, supra note 3 at 241.

See Lombard et al., supra note 200.

See supra text at notes ____.
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 0.81-1.00 is Almost Perfect. 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 either 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 re-testing 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.

4. Alternative Coding Techniques

Researches 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. This technique may occur to legal researchers because it mimics the familiar 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. 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 have coders confer when they disagree in order to seek consensus, or the researcher uses her own expertise to resolved disagreements. Again, this may or may not improve reliability, but it does not establish reliability. The process of reaching consensus might be arbitrary, or the author’s expert view may not be objectively reproducible.

A variation of these techniques is expert panel consensus. 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. This has been shown to be a

221 Also illustrating this technique, see Johnson, supra note 144 at 184-85 (comparing agreement rates for 2-category versus 4-category coding).
222 Phelps & Gates, supra note 144; Johnson, supra note 117.
223 See Krippendorf, supra note 3 at 217.
224 Lawlor, supra note 101 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.”
fairly reliable method for rating highly complex and judgmental aspects of medical decisionmaking.\textsuperscript{226} 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.\textsuperscript{227} Also noteworthy are researchers who analyze judicial texts entirely by computer, looking for revealing patterns in syntax or semantics.\textsuperscript{228} 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.\textsuperscript{229}

C. Statistical Analysis

A credible content analyst does not necessarily 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 geographic exploration. Often, simple counts and percents are sufficient to document more rigorously an author’s claim about trends in the caselaw, to challenge conventional wisdom, or to raise provocative questions meriting further study. Because many case-counting studies code the entire universe of relevant cases,\textsuperscript{230} statistics are not essential for analyzing the probability that the sample cases reflect the reality in a larger population.\textsuperscript{231}


\textsuperscript{227} Groscup et al., \textit{supra} note 138; Wahlbeck, \textit{supra} note 117.


\textsuperscript{230} See \textit{supra}, text at notes

\textsuperscript{231} 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.
Moreover, content analysis need not involve numbers at all. Instead, it can employ rigorous methods of purely qualitative analysis\textsuperscript{232} that focus on themes and patterns that are best described through conceptual description and narrative illustrations rather than numbers.\textsuperscript{233}

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 another. Thirty percent (28/95) of 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.”\textsuperscript{234} 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.\textsuperscript{235}

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 most important predictors of decisions. (Sometimes, factors that legal analysts thought were dominant turn out to be red herrings.)\textsuperscript{236} Alternatively, factors that, standing alone, may not appear significant might emerge as such once the influence of other factors is controlled for statistically.\textsuperscript{237}

\textsuperscript{232} Often, researchers use computer programs such as Nudist and Ethnograph to assist with this analysis. See generally COLLECTING AND INTERPRETING QUALITATIVE MATERIALS (N. K. Denzin & Y. S. Lincoln eds.) (2003); M.Q. PATTON, QUALITATIVE RESEARCH AND EVALUATION METHODS (3rd ed. 2002).

\textsuperscript{233} 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 69.

\textsuperscript{234} Or, more technically, this is known as the multiple comparison problem, causing a false discovery rate. Or, more colloquially, one of us once heard it described as “dancing barefoot through a field of data” without any particular goal in mind.

\textsuperscript{235} 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 100 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 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.

\textsuperscript{236} See text at notes ___ supra.

\textsuperscript{237} See text at notes ___ supra.
Regression analysis is especially well suited for finding hidden patterns and associations in a complex body of caselaw in which legal analyses or results depend “extensively on discrete and numerous factors, occurring in varying combinations in the cases” which courts weigh “without apparent fixed standards.”\textsuperscript{238} 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.\textsuperscript{239}

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) in 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.\textsuperscript{240} 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)\textsuperscript{241} 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 adopting or rejecting legal rules across 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 — two distinct ways of talking about and coming to know what judges do in legal opinions.

It is a strength of the academy 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 caselaw conventionally, as precedent-establishing legal doctrine. We found only a few examples, but they were instructive. One method analyzes caselaw using jurisdictions, rather than cases, as the unit of analysis.\textsuperscript{242} 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.\textsuperscript{243} 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.\textsuperscript{244} A difficulty with this approach, however, is deciding how much weight to

\begin{itemize}
  \item \textsuperscript{238} Haar et al., \textit{supra} note 7 at 656.
  \item \textsuperscript{239} McChesney, \textit{supra} note 75.
  \item \textsuperscript{240} Hammer & Sage, \textit{supra} note 27.
  \item \textsuperscript{241} Phelps & Gates, \textit{supra} note 144, for instance, uses each paragraph of an opinion as the unit of analysis.
  \item \textsuperscript{242} \textit{E.g.}, Morriss, \textit{supra} note 47.
  \item \textsuperscript{243} \textit{E.g.}, KLEIN, \textit{supra} note 117.
  \item \textsuperscript{244} See LANDES AND POSNER, \textsc{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
Without independent, objective verification of how much weight 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 amount of case coding work that has been done to date, much development and innovation lie ahead in order to find the best uses for this most quintessentially legal means of empirical study.

IV. 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 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 caselaw 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. Basic components such as replicable case selection and coded reading can be done profitably by any legal analyst. 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, for instance, who study mainly the outcome of cases and their subject matter,
or 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 caselaw, 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 caselaw empiricism simply because the relevant phenomena for lawyers consist of texts: written opinions.

Content analysis is much more, though, than a better way to read cases. 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 will be treated seriously in the rest of the social science world. 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 an empirical methodology that is uniquely legal.