



University of Arkansas at Little Rock William H. Bowen School of Law
Bowen Law Repository: Scholarship & Archives

Faculty Scholarship

2016

Using the West Key Number System as a Data Collection and Coding Device for Empirical Legal Scholarship: Demonstrating the Method Via a Study of Contract Interpretation

Joshua M. Silverstein

University of Arkansas at Little Rock William H. Bowen School of Law, jmsilverstei@ualr.edu

Follow this and additional works at: http://lawrepository.ualr.edu/faculty_scholarship



Part of the [Contracts Commons](#), and the [Legal Writing and Research Commons](#)

Recommended Citation

Joshua M. Silverstein, Using the West Key Number System as a Data Collection and Coding Device for Empirical Legal Scholarship: Demonstrating the Method Via a Study of Contract Interpretation, 34 J.L. & Com. 203 (2016).

This Article is brought to you for free and open access by Bowen Law Repository: Scholarship & Archives. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of Bowen Law Repository: Scholarship & Archives. For more information, please contact mmserfass@ualr.edu.

ARTICLES

USING THE WEST KEY NUMBER SYSTEM AS A DATA COLLECTION AND CODING DEVICE FOR EMPIRICAL LEGAL SCHOLARSHIP: DEMONSTRATING THE METHOD VIA A STUDY OF CONTRACT INTERPRETATION

*Joshua M. Silverstein**

Empirical research is an increasingly important type of legal scholarship. Such research generally requires the collection and coding of large quantities of data. These tasks pose critical challenges for legal scholars. Most crucially, they are often resource-intensive. The primary purpose of this article is to explain how researchers can use the West Key

* Professor of Law, University of Arkansas at Little Rock, William H. Bowen School of Law. B.A., *summa cum laude*, 1993, Hamilton College. J.D., *magna cum laude*, 1996, New York University School of Law. Completion of this article was made possible by a grant from the University of Arkansas at Little Rock, William H. Bowen School of Law.

I would like to thank Jasmine Abdel-khalik, Javitt Adili, Ian Ayres, Royce Barondes, Terri Beiner, Ken Gallant, Mark Gergen, Sarah Hobbs, Nick Kahn-Fogel, Russell Korobkin, Kim Love-Myers, Michael Hunter Schwartz, Phyllis Silverstein, Robert Scott, and Robert Steinbuch for their comments on drafts of this Article. I would also like to thank Kim Love-Myers and Fei Liu of the University of Georgia Statistical Consulting Center for their work in conducting the statistical analysis set forth in Appendix 2. In addition, my thanks to the various Thomson Reuters reference attorneys and other staff members who helped me obtain answers to multiple questions regarding the Key Number System and several related Thomson Reuters products.

I presented this Article to the faculty at the University of Arkansas at Little Rock, William H. Bowen School of Law. My thanks to all those who attended and offered comments. And thanks to the Faculty Development Committee of the Bowen Law School for organizing the presentation. I also presented this article at the 10th International Conference on Contracts held at the University of Nevada Law Vegas, William S. Boyd School of Law. My thanks to all those who attended and offered comments. Finally, I would like to thank Nikki Cox, Mary Erhart, Kyla Farler, Daniel Ford, Lauren Guinn, Ryan Lazenby, and Whitney Ohlhausen for their invaluable research assistance in preparing this article.

Number System to dramatically streamline the process of data collection and coding. The article accomplishes this, in part, through a demonstration: it employs the Key Number System to conduct an empirical study of contract interpretation.

Contract interpretation is one of the most significant areas of commercial law. And the subject has received considerable scholarly attention during the last decade. Virtually all academic work in this field is doctrinal or theoretical. But numerous contract interpretation issues cry out for empirical investigation. The secondary purpose of this article is to test one of the central claims in the judicial and academic debate over the optimal method of contract interpretation—the claim that the “contextualist” approach to interpretation results in more litigation over the meaning of contracts than does the “textualist” approach. The results of the study set forth below are inconsistent with that thesis. By thirteen of fourteen measures, there was no statistically significant difference in the amount of interpretation litigation between textualist and contextualist regimes. And for the fourteenth measure, while there was a statistically significant difference, the result was the opposite of that predicted by textualist theory: there was more litigation under textualism.

TABLE OF CONTENTS

I.	Introduction.....	206
II.	Data Collection and Coding Challenges in Empirical Legal Research.....	209
III.	Using the West Key Number System as a Data Collection and Coding Device.....	216
	A. The West American Digest System.....	216
	B. Literature Review of Empirical Research Employing the Key Number System.....	220
	C. Methodological Issues.....	226
IV.	Contract Interpretation.....	253
	A. The Textualist and Contextualist Approaches.....	253
	B. The Policy Debate.....	261
	1. Interpretive Accuracy.....	263
	2. Transaction Costs.....	268
	3. Enforcement Costs.....	270
	4. Other Policy Arguments.....	278
	5. The Need for Empirical Evidence.....	283
V.	An Empirical Study of the Impact of the Textualist and Contextualist Approaches on Enforcement Costs.....	284
	A. Study Outline.....	284
	B. Study Methodology, Raw Totals, and Statistical Analysis.....	286
	C. Discussion.....	300
	1. Methodological Explanations for the Study Findings.....	300
	2. Alternative Explanations for the Study Findings.....	305
IV.	Conclusion.....	308
	Appendix 1: Westlaw Searches Used in the Study.....	313
	Appendix 2: Statistical Analysis.....	317
	A. t-tests.....	317
	B. Logistic Regressions.....	321

I. INTRODUCTION

Numerous legal and policy issues involve questions of empirical fact. But legal academics have historically eschewed empirical research.¹ As a result, many of the normative recommendations contained in law review articles are based on little more than informed speculation.²

Over the last two decades, legal scholarship has changed. Empirical research on law is an increasingly common and important type of academic work.³ This is certainly true in the field of contracts, where some of the theories advanced by contract scholars are finally receiving the type of empirical scrutiny necessary to test their validity.⁴ “There is widespread

¹ Tracey E. George, *An Empirical Study of Empirical Legal Scholarship: The Top Law Schools*, 81 IND. L.J. 141, 141 (2006) (explaining that empirical scholarship “was uncommon in law schools through most of the last century”); Peter H. Schuck, *Why Don't Law Professors Do More Empirical Research?*, 39 J. LEGAL EDUC. 323, 323 (1989); *see also id.* at 331–33 (offering nine explanations for the paucity of empirical work, including law professors’ lack of training in empirical methods, the tedious nature of the work, and the resources required); Craig Allen Nard, *Empirical Legal Scholarship: Reestablishing a Dialogue Between the Academy and Profession*, 30 WAKE FOREST L. REV. 347, 361–65 (1995) (reaching similar conclusions based on a survey of law professors).

² *See* Edward L. Rubin, *The Practice and Discourse of Legal Scholarship*, 86 MICH. L. REV. 1835, 1896 (1988) (“In discussing the effects of a judicial decision, for example, legal scholars tend to opt for speculations that are structured by their own legal discourse, rather than searching for a methodology that can be used to determine these effects on empirical grounds.”); *see also* Elizabeth Warren & Jaw Lawrence Westbrook, *Searching for Reorganization Realities*, 72 WASH. U. L.Q. 1257, 1261–63, 1286–89 (1994) (explaining that legal scholarship that addresses bankruptcy policy suffers from a “dominance of speculation over systemic empirical inquiry” and calling for scholars to both formulate more empirically-testable hypotheses and produce more empirical studies).

³ George, *supra* note 1, at 142 (“But [empirical legal scholarship] recently and dramatically has expanded in law reviews, at conferences, and among leading law faculties.”) (collecting authorities); Julie Graves Krishnaswami, *Nicholas Freudenberg: A Selected Bibliography to Accompany a Conversation on Health and Law*, 12 N.Y. CITY L. REV. 55, 57 (2008) (“[E]mpirical methods are becoming important within legal scholarship.”); Carolyn Shapiro, *Coding Complexity: Bringing Law to the Empirical Analysis of the Supreme Court*, 60 HASTINGS L.J. 477, 477–78 (2009) (“The legal academy has recently experienced a surge of interest in quantitative empirical analysis.”); *see generally* Symposium, *The Empirical Revolution in Law*, 65 STAN. L. REV. 1195 (2013); Symposium, *Empirical and Experimental Methods in Law*, 2002 U. ILL. L. REV. 791 (2002). For an empirical study of the increase in one type of empirical legal scholarship, *see* Mark A. Hall & Ronald F. Wright, *Systematic Content Analysis of Judicial Opinions*, 96 CALIF. L. REV. 63, 70–73 (2008).

⁴ *See* Russell Korobkin, *Empirical Scholarship in Contract Law: Possibilities and Pitfalls*, 2002 U. ILL. L. REV. 1033 (2002) (setting forth an extensive overview of empirical legal scholarship in the field of contracts); George S. Geis, *Automating Contract Law*, 83 N.Y.U. L. REV. 450, 454 n.18 (2008) (collecting similar authorities published after Korobkin’s article); *see also* Zev J. Eigen, *Empirical Studies of Contract*, 8 ANN. REV. L. & SOC. SCI. 291 (2012) (presenting an overview of empirical work

belief that empirical research can improve both our positive understanding of law and behavior and inform our normative views of legal doctrine.”⁵ The increase in empirical scholarship is thus a welcome development.

However, empirical research poses critical challenges.⁶ For example, legal academics generally lack training in quantitative methods, a fact that is cited by many when criticizing the quality of empirical scholarship that appears in law reviews.⁷

Other challenges concern the collection and coding of data.⁸ Most significantly, gathering and organizing legal materials is often time-consuming and expensive. The primary purpose of this article is to explain how researchers can use the West Key Number System to dramatically streamline the process of data collection and coding. The article accomplishes this, in part, by using the Key Number System to conduct an empirical study of contract interpretation.

Contract interpretation is one of the most important topics in commercial law.⁹ And the subject has received a great deal of academic attention during the last decade.¹⁰ Indeed, interpretation is recognized as

on contracts in law and several other disciplines). Prior to the recent, general upsurge in empirical work, contracts suffered from the same lack of empirical research as other legal fields. See Russell J. Weintraub, *A Survey of Contract Practice and Policy*, 1992 WIS. L. REV. 1, 4 (1992) (“Despite [the] need for data . . . to date there have been only a handful of empirical studies focusing on particular contract problems and relationships . . .”).

⁵ Korobkin, *supra* note 4, at 1050 & n.97 (collecting authorities); see also Geis, *supra* note 4, at 452-53 n.12, 454 n.17 (collecting authorities calling for an increase in empirical research).

⁶ See Geis, *supra* note 4, at 453-54; Michael Heise, *The Importance of Being Empirical*, 26 PEPP. L. REV. 807, 810-24 (1999).

⁷ See, e.g., Lee Epstein & Gary King, *The Rules of Inference*, 69 U. CHI. L. REV. 1, 6 (2002) (explaining that “the current state of empirical legal scholarship is deeply flawed”); *id.* at 9 (“One source of the problem almost certainly lies in the training law professors receive.”). The lack of training is also frequently identified as a reason that legal academics conducted little empirical work prior to the last two decades. See, e.g., Heise, *supra* note 6, at 817-18; Schuck, *supra* note 1, at 333.

⁸ See *infra* Part II.

⁹ See STEVEN J. BURTON, ELEMENTS OF CONTRACT INTERPRETATION 1 (2009) (“Issues of contract interpretation are important in American law. They probably are the most frequently litigated issues on the civil side of the judicial docket.”); Benjamin E. Hermalin et al., *Contract Law*, in 1 HANDBOOK OF LAW & ECONOMICS § 4, at 68 (A. Mitchell Polinsky & Steven Shavell eds., 1st ed. 2007) (“Probably the most common source of contractual disputes is differences in interpretation . . .”); Alan Schwartz & Robert E. Scott, *Contract Interpretation Redux*, 119 YALE L.J. 926, 928 & n.3 (2010) (“[C]ontract interpretation remains the largest single source of contract litigation between business firms.”) (collecting authorities).

¹⁰ Steven J. Burton, *A Lesson on Some Limits of Economic Analysis: Schwartz and Scott on Contract Interpretation*, 88 IND. L.J. 339, 340 (2013) (“After decades of relative neglect, contract

“the least settled, most contentious area of contemporary contract doctrine and scholarship.”¹¹ The vast majority of the work in this field is theoretical.¹² But numerous contract interpretation issues cry out for empirical investigation.¹³ The secondary purpose of this article is to test one of the central claims in the debate between the “textualist” and “contextualist” schools of interpretation.

Textualists argue that the interpretation of contracts should focus primarily on the language contained within the four corners of the parties’ agreement. Contextualists believe that interpreting courts generally ought to consider both the language of the parties’ contract and extrinsic evidence.¹⁴ One of the signature claims advanced by textualists in support of their position is that the contextualist approach results in more litigation over the meaning of contracts than the textualist approach.¹⁵ Critically, no commentator or court has proffered empirical evidence addressing this contention. My article is a first step towards filling that gap in the secondary literature and caselaw.

The results of the study set forth below are inconsistent with the textualist thesis. By thirteen of fourteen measures, there was no statistically significant difference in the amount of interpretation litigation between textualist and contextualist regimes. And for the fourteenth measure, while there was a statistically significant difference, the result was the opposite of that predicted by textualist theory: there was more litigation under *textualism*.

The data for this study was compiled solely via Westlaw’s electronic research tools. Most importantly, all of the coding was completed by West through its Key Number System. I did not manually code any of the decisions in my dataset. To construct my coded dataset of 8,113 cases, I simply identified the correct Westlaw search queries. Employing West’s

interpretation became a hot topic of scholarly debate after 2003.”); *id.* at 340 n.8 (collecting authorities); David McLaughlan, *Contract Interpretation: What Is It About?*, 31 SYDNEY L. REV. 5, 5 (2009) (“In recent times, contract interpretation has become one of the most contentious areas of the law of contract.”).

¹¹ Ronald J. Gilson et al., *Text and Context: Contract Interpretation as Contract Design*, 100 CORNELL L. REV. 23, 25 (2014).

¹² See *infra* Part IV.B.5.

¹³ See *infra* Part IV.B.5.

¹⁴ See *infra* Part IV.A.

¹⁵ See *infra* Part IV.B.3.

Key Number System in this manner dramatically sped up my work and made the cost of the study negligible.

Of course, there are limits to the usefulness of this methodology.¹⁶ But despite those limits, I believe that my approach can serve as a valuable tool in empirical legal research by making data collection and coding significantly easier and cheaper in many circumstances.

Part II of this article discusses the challenges scholars face in collecting and coding data for empirical work. Part III begins with an overview of the West American Digest System, of which the Key Number System is a part. It next summarizes how scholars have employed the Key Number System in empirical legal scholarship. Part III ends with a discussion of the methodological issues surrounding the use of the Key Number System as a collection and coding device. Part IV describes the textualist and contextualist approaches to contract interpretation and the policy justifications underlying each approach. Part V sets forth the specific methodology and results of my study and then discusses the findings. Finally, Part VI contains some concluding thoughts.¹⁷

II. DATA COLLECTION AND CODING CHALLENGES IN EMPIRICAL LEGAL RESEARCH

The process of collecting and coding data lies at the heart of empirical research.¹⁸ Empirical data in legal studies is typically gathered via (1) surveys and interviews, (2) experiments, and (3) the review of material stored in archives or databases.¹⁹ Once the data is collected, it must be coded. Coding is the ““process of transforming raw data into standardized

¹⁶ See, e.g., *infra* Part III.C.

¹⁷ There are also two appendices. Appendix 1 sets forth the Westlaw searches I ran for the study and Appendix 2 contains the statistical analysis of the data recovered by the searches.

¹⁸ See Will Rhee, *Evidence-Based Federal Civil Rulemaking: A New Contemporaneous Case Coding Rule*, 33 PACE L. REV. 60, 99–100 (2013) (“Generally, the methodology for all empirical research has four steps: (1) design the empirical project; (2) collect and code data; (3) analyze the data; and (4) present the final results Whether qualitative, quantitative, experimental, or multi-method, all empirical research must code raw data into standardized variables that can be analyzed.”).

¹⁹ See ROBERT M. LAWLESS ET AL., *EMPIRICAL METHODS IN LAW* 59–138 (2010); see also Korobkin, *supra* note 4, at 1034 (dividing empirical contract law scholarship into four categories based on the source of empirical data—“judicial opinions, actual contracting practices, contracting experiments, and nonexpert opinion about contract law”).

form”²⁰ so that it can be systematically analyzed for research purposes,²¹ such as via statistical techniques.²²

Data collection and coding both present important challenges. For example, researchers must address numerous methodological issues in designing a survey,²³ an experiment,²⁴ or an archival study.²⁵ And coding the information obtained using these techniques raises additional methodological questions.²⁶

The challenge of greatest concern for my purposes here is that the process of collecting and coding data is usually time-consuming and expensive.²⁷ Indeed, additional funding is often required²⁸ so that scholars can (1) hire and train support staff needed to gather and review data,

²⁰ LAWLESS ET AL., *supra* note 19, at 166 (quoting EARL R. BABBIE, *THE PRACTICE OF SOCIAL RESEARCH* 325 (11th ed. 2006)) (further explaining that “systematizing or standardizing the data . . . captures the essence of what it means to code data”).

²¹ *Id.* (citing Lee Epstein & Andrew Martin, *Coding Variables*, in *ENCYCLOPEDIA OF SOCIAL MEASUREMENT* (Kimberly Kempf-Leonard ed., 2005)).

²² *Id.*

²³ *Id.* at 86 (“Nevertheless, to design an effective survey . . . , it is important to bear in mind the principles of survey design that we have articulated in this chapter.”); *see generally id.* at 59–86 (discussing survey methodologies).

²⁴ *Id.* at 122 (“Whether an experiment is conducted in the laboratory or in the field, careful attention to the setting up of conditions that are suited to testing a particular hypothesis is critical.”); *see generally id.* at 93–122 (discussing methods of conducting experiments).

²⁵ *Id.* at 136 (noting that researchers must “pay close attention to the nature of the materials . . . under consideration, to the ways in which the data have been abstracted from them, and to their relevance to the research questions of interest”); *see generally id.* at 125–36 (discussing empirical research based on archival or database materials).

²⁶ *See generally id.* at 165–83 (discussing proper coding practices); *id.* at 128 (further explaining the process of coding archival or database materials).

²⁷ LAWLESS ET AL., *supra* note 19, at 128 (observing that coding archival material “can be very time consuming”); Tracy E. George, *An Empirical Study of Empirical Legal Scholarship: The Top Law Schools*, 81 *IND. L.J.* 141, 151 (2006) (noting that data collection is “time-intensive”); Michael Heise, *The Past, Present, and Future of Empirical Legal Scholarship: Judicial Decision Making and the New Empiricism*, 2002 *U. ILL. L. REV.* 819, 829 (2002) (“Unfortunately, data gathering is frequently labor-intensive and time-consuming and, consequently, often quite expensive.”); Korobkin, *supra* note 4, at 1050 (“[E]mpirical research typically is more expensive and time consuming than purely theoretical work or traditional doctrinal analysis . . .”).

²⁸ Numerous empirical studies that I came across in writing this article noted that they were made possible by funding support from sources outside the author’s institution. *See, e.g.*, Timothy R. Johnson et al., *Oral Advocacy Before the United States Supreme Court: Does It Affect the Justices’ Decisions?*, 85 *WASH. U. L. REV.* 457, 457 n.d1 (2007); Herbert M. Kritzer, *The Commodification of Insurance Defense Practice*, 59 *VAND. L. REV.* 2053, 2053 n.a1 (2006).

(2) obtain access to archives, databases, and other materials, and (3) pay research subjects.

Let me elaborate with respect to data coding. For empirical work to be effective, information must be coded in a reliable manner. This means that “the same data are coded in the same way at different times or by different coders.”²⁹ To ensure this level of standardization, any individuals hired to perform coding must be thoroughly trained and provided with detailed codebooks to aid them in their work.³⁰ It is also recommended that at least two people code every piece of data and that researchers compare whether the coders recorded their results consistently.³¹

Good [coding] practices almost always involve an iterative process, where the researcher must code some observations, assess the coding reliability, revise the coding procedures, and then begin the process again. Reliable coding often requires multiple passes through the data, coding different aspects of the data or more finely tuning particular pieces of information.³²

Such protocols normally requires considerable time and money. But failing to follow them fundamentally undercuts the validity of one’s research.³³

To further illustrate the time and resources required to collect and code data for empirical work, I have set forth below summaries of several studies from the contracts literature. While I included studies that reflect each of the three methods of data gathering—surveys, experiments, and archival research—the bulk of the examples come from the third category since that is the type of scholarship to which my project is most comparable.³⁴

Stewart Macaulay’s groundbreaking article on the contracting practices of business people was built around a survey. Macaulay interviewed sixty-eight business persons and lawyers, with the meetings ranging in length from thirty minutes to six hours.³⁵ After each interview,

²⁹ LAWLESS ET AL., *supra* note 19, at 166.

³⁰ *Id.* at 180.

³¹ *Id.* at 181. This is known as “double coding” or “double data entry.” *Id.*

³² *Id.* at 168; accord Hall & Wright, *supra* note 3, at 80–81, 107–17 (describing the protocols for the systematic coding of judicial opinions).

³³ See LAWLESS ET AL., *supra* note 19, at 166–68 (setting forth an example).

³⁴ See *id.* at 125–38 (explaining archival and database research). Note that my study might also be classified as a quasi-experiment, see *id.* at 115–17 (explaining quasi-experiments), and thus placed into the second category (experiments) rather than the third.

³⁵ Stewart Macaulay, *Non-Contractual Relations in Business: A Preliminary Study*, 28 AM. SOC. REV. 55, 55 (1964).

he dictated a detailed report of the discussion.³⁶ Based on the information obtained, Macaulay concluded that contract law appears to have a small effect on how business people organize their dealings and that most individuals who participate in the contracting process have little knowledge of contract law.³⁷ While Macaulay's study was invaluable, the data-gathering process he used was quite time-consuming.

More recently, Yair Listokin conducted an experiment in which identical used iPods were sold via auction on eBay with randomly varying warranties.³⁸ Consumers paid different prices based on the type of warranty, and the precise spread in the sales prices suggested that the buyers correctly understood the scope of their legal protection when there was no express warranty.³⁹ This research also took a great deal of time to conduct. In fact, the 166 iPods sold as part of the study were purchased on eBay over a thirteen-month period.⁴⁰

Legal scholarship that fits into the third category (archive and database research) often analyzes judicial opinions. For example, in one study, Daniel Farber and John Matheson read every published opinion that cited section 90 of the first or second restatement of contracts over a ten-year period—a total of 222 cases.⁴¹ From this, they drew conclusions about which factors determine whether a court will find liability under the theory

³⁶ *Id.*

³⁷ *See id.* at 56–61. For another example, see Weintraub, *supra* note 4, at 2 (survey of eight-four corporate general counsel about a range of contracting practices). For an example from outside the field of contracts, see Mark D. Gough, *The High Costs of an Inexpensive Forum: An Empirical Analysis of Employment Discrimination Claims Heard in Arbitration and Civil Litigation*, 35 BERKELEY J. EMP. & LAB. L. 91, 91 (2014) (survey of approximately 700 plaintiff employment lawyers; finding that “employee win rates and award amounts are lower in arbitration compared to those found in federal and state court” and that the divergence cannot be explained by “systematic differences in case characteristics between the forums”).

³⁸ Yair Listokin, *The Meaning of Contractual Silence: A Field Experiment*, 2 J. LEGAL ANALYSIS 397, 397, 410 (2010).

³⁹ *Id.* at 397, 410.

⁴⁰ *Id.* at 402. For another example, see Ian Ayres, *Further Evidence of Discrimination in New Car Negotiations and Estimates of Its Cause*, 94 MICH. L. REV. 109, 113–15, 119 (1995) (experiment in which shells were trained to negotiate over the price of cars with actual salespeople using the same script; the shells varied by race and sex, enabling Professor Ayres to draw conclusions about the effects of these traits on the negotiating behavior of the salespeople); *see also* Ian Ayres, *Fair Driving: Gender and Race Discrimination in Retail Car Negotiations*, 104 HARV. L. REV. 817 (1991).

⁴¹ Daniel A. Farber & John H. Matheson, *Beyond Promissory Estoppel: Contract Law and the “Invisible Handshake,”* 52 U. CHI. L. REV. 903, 907 n.14 (1985).

of promissory estoppel.⁴² In particular, they deduced that the presence or absence of reliance does not explain the results in promissory estoppel cases despite reliance's status as an element of the cause of action.⁴³ Robert Hillman conducted a similar study thirteen years later. Hillman began by using electronic research tools to compile a data set of every reported decision over a two-year period in which a promissory estoppel claim succeeded or failed or in which the cause of action was discussed⁴⁴—a total of 362 cases.⁴⁵ He then “devised a coding sheet that asked fifty-two questions about each case,” which his research assistants used to read and code the decisions.⁴⁶ Hillman's results contradicted those of Farber and Matheson: He concluded that reliance plays a critical role in determining whether a promissory estoppel claim is successful.⁴⁷ Note that what distinguishes studies like those by Farber, Matheson, and Hillman from traditional doctrinal scholarship is “that they attempt to use objective facts found in judicial opinions as data points in their analysis, rather than examining the internal logic or reasoning of opinions.”⁴⁸

Other studies that fit within the third category focus on the terms contained in contracts. To illustrate, Daniel Schwarcz obtained from state insurance authorities the homeowners policies offered by twenty-four insurance groups across six states.⁴⁹ Schwarcz then analyzed the provisions of each policy.⁵⁰ He found that homeowners policies vary significantly from carrier to carrier, undercutting the common claim that homeowners

⁴² *Id.* at 925–29.

⁴³ *Id.* at 904, 910.

⁴⁴ Robert A. Hillman, *Questioning the “New Consensus” on Promissory Estoppel: An Empirical and Theoretical Study*, 98 COLUM. L. REV. 580, 582 (1998).

⁴⁵ *Id.* at 582–83 n.15.

⁴⁶ *Id.* at 582–83.

⁴⁷ *Id.* at 580, 583, 619 (also concluding that promissory estoppel claims have a “remarkable lack of success . . . in reported decisions”).

⁴⁸ Korobkin, *supra* note 4, at 1040; accord Hall & Wright, *supra* note 3, at 64–66, 76 (explaining the difference between “systematic content analysis” and traditional doctrinal scholarship) (“Using this method [of systematic content analysis], a scholar collects a set of documents, such as judicial opinions on a particular subject, and systematically reads them, recording consistent features of each and drawing inferences about their use and meaning.”).

⁴⁹ Daniel Schwarcz, *Reevaluating Standardized Insurance Policies*, 78 U. CHI. L. REV. 1263, 1277–78 (2011).

⁵⁰ *Id.* at 1279.

insurance is highly standardized.⁵¹ Florencia Marotta-Wurgler’s study of end-user license agreements (“EULAs”) for software products is another helpful example.⁵² Marotta-Wurgler examined 647 EULAs from 598 different software manufacturers after obtaining the agreements from the companies’ websites or by direct request.⁵³ She assessed each contract using an index of “buyer friendliness” that was based on how the agreements handled “23 important and common terms that allocate rights between buyers and sellers of software.”⁵⁴ Among other conclusions, Marotta-Wurgler found that EULAs almost uniformly favor the seller more than the default rules set forth in the Uniform Commercial Code.⁵⁵

The previous two paragraphs summarized four archival or database studies. Two of these studies—those conducted by Schwarcz and Marotta-Wurgler on the terms of actual contracts—involved significant data *collection* efforts. But all four projects required extensive data *coding*. In particular, for each study, the authors or their research assistants had to read and code every case or agreement in the dataset. This is standard practice in archival empirical work, whether the inquiry concerns contracts⁵⁶ or other legal subjects.⁵⁷

⁵¹ *Id.* at 1263.

⁵² Florencia Marotta-Wurgler, *What’s in a Standard Form Contract? An Empirical Analysis of Software License Agreements*, 4 J. EMPIRICAL LEGAL STUD. 677 (2007).

⁵³ *Id.* at 678–79, 681.

⁵⁴ *Id.* at 679.

⁵⁵ *Id.* at 680.

⁵⁶ See, e.g., Nathan M. Crystal, *An Empirical View of Relational Contracts Under Article 2 of the Uniform Commercial Code*, 1988 ANN. SURV. AM. L. 293, 299 (1988) (empirical study of the frequency with which relational contracts appear in litigation under Article 2 of the Uniform Commercial Code); Larry A. DiMatteo & Bruce Louis Rich, *A Consent Theory of Unconscionability: An Empirical Study of Law in Action*, 33 FLA. ST. U. L. REV. 1067, 1091–94 (2006) (empirical study of unconscionability law); Jeffrey L. Harrison, *Trends and Traces: A Preliminary Evaluation of Economic Analysis in Contract Law*, 1988 ANN. SURV. AM. L. 73, 78–79, 80 (1988) (empirical study of the influence of law and economics scholarship on contract law); Fred S. McChesney, *Tortious Interference with Contract Versus “Efficient” Breach: Theory and Empirical Evidence*, 28 J. LEGAL STUD. 131, 173–76 (1999) (empirical study of the circumstances in which the law recognizes a claim for tortious interference with contract); James J. White, *Evaluating Article 2 of the Uniform Commercial Code: A Preliminary Empirical Expedition*, 75 MICH. L. REV. 1262, 1264, 1272, 1274–85 (1977) (empirical study of litigation under Article 2 of the Uniform Commercial Code).

⁵⁷ For some classic examples, see Robert A. Kagan et al., *The Business of State Supreme Courts, 1870–1970*, 30 STAN. L. REV. 121, 122, 124–28 (1977) (empirical study of the “business” of state supreme courts—“the volume of cases handled, the type of issues that appear, the kind of litigants served, the rate of affirmance or reversal of lower court decisions, and the variation of these factors

For many empirical studies, data collection and coding will necessarily take significant time and resources. But in some circumstances, I believe that the West Key Number System can be used to dramatically streamline both aspects of the process, especially the coding of data. That system is discussed in the next part.

across states and over time”); the authors read and coded 5,904 cases obtained using a variety of selection criteria); Thomas W. Merrill, *Judicial Deference to Executive Precedent*, 101 YALE L.J. 969, 980–82 (1992) (empirical study of the United States Supreme Court’s jurisprudence regarding judicial deference to administrative agency interpretations of statutes; the author read and coded every Supreme Court case on this subject from a ten-year time period, a total of ninety cases); Richard A. Posner, *A Theory of Negligence*, 1 J. LEGAL STUD. 29, 29, 34–35 (1972) (empirical study of the nature negligence law; the author read and coded every published appellate decision involving an accident issued in the first quarter of 1875, 1885, 1895, and 1905, a total of 1,528 cases); Vicki Schultz, *Telling Stories About Women and Work: Judicial Interpretations of Sex Segregation in the Workplace in Title VII Cases Raising the Lack of Interest Argument*, 103 HARV. L. REV. 1749, 1750, 1766–67 (1990) (empirical study of a defense used by employers to challenge employment discrimination claims based on sex; the author read and coded fifty-four cases, which constituted every published decision in which a lower federal court addressed the defense at issue).

For some more recent examples, see Barton Beebe, *An Empirical Study of the Multifactor Test for Trademark Infringement*, 94 CALIF. L. REV. 1581, 1581, 1649–50 (2006) (empirical study of the multifactor tests used to assess the likelihood of consumer confusion in trademark law; the author read and coded all reported federal district court decisions for a five-year period in which a multifactor test was used, a total of 331 opinions; the dataset was created by reviewing 1,252 opinions identified via Westlaw and Lexis and eliminating the 921 cases that did not make “substantial use” of a multifactor test); Stephen F. Befort, *An Empirical Examination of Case Outcomes Under the ADA Amendments Act*, 70 WASH. & LEE L. REV. 2027, 2045–49 (2013) (empirical study of the impact of an amendment to the Americans With Disabilities Act; the author read and coded 237 reported federal court summary judgment decisions; the dataset was created by examining 1,289 reported opinions identified in a Westlaw search and eliminating the 1,052 cases that failed to meet certain criteria); Patricia W. Hatamyar, *The Tao of Pleading: Do Twombly and Iqbal Matter Empirically?*, 59 AM. U. L. REV. 553, 585–86, 589–95 (2010) (empirical study of the impact of the change in federal pleading standards; the author read and coded 1039 cases obtained via Westlaw); Nick Kahn-Fogel, *Manston and Its Progeny: An Empirical Analysis of American Eyewitness Law*, 3 ALA. C.R. & C.L. L. REV. 175, 177–78, 196–209 (2012) (empirical study of the degree to which judicial evaluation of eyewitness identification testimony is consistent with the science of eyewitness identification; the author read and coded 1,471 cases available on Westlaw that cited to the United States Supreme Court’s seminal eyewitness identification decision).

For an excellent survey of legal scholarship engaging in the systematic coding of cases, see Hall & Wright, *supra* note 3, at 67–76, 82 n.72, 90–95, 101–02; *see also* Geis, *supra* note 4, at 479 (explaining that the empirical study of caselaw “typically requires intensive tagging, coding, and interpretation”).

III. USING THE WEST KEY NUMBER SYSTEM AS A DATA COLLECTION AND CODING DEVICE

A. *The West American Digest System*

Digests are tools for locating caselaw on a particular subject.⁵⁸ “[E]ditors of commercial publishers analyze opinions and write brief descriptive abstracts of the various points of law in the opinions.”⁵⁹ The abstracts are often called “headnotes” or “digest paragraphs”⁶⁰ and they are typically included in the commercial reporters prior to the official text of each decision.⁶¹ The headnotes are also “arranged by subject in classification systems and published in sets known as digests.”⁶² A digest operates like an index; “instead of simple one-line entries, however, it consists of paragraphs describing the legal principles decided.”⁶³

Digests vary by the types of cases they include.⁶⁴ Some contain decisions from only a single state or court system.⁶⁵ Others focus on a small group of states or on a particular legal subject.⁶⁶ The West American Digest System is intended to be a comprehensive digest, serving as a “master index” to all of American caselaw.⁶⁷

West classifies headnotes in the American Digest via the “West Key Number System.”⁶⁸ “The West Key Number analysis of American law is a scheme whereby the entire body of American law has been conceptualized and arranged in an orderly and logical way.”⁶⁹ The Key Number System

⁵⁸ STEVEN M. BARKAN ET AL., *FUNDAMENTALS OF LEGAL RESEARCH* 78 (9th ed. 2009).

⁵⁹ *Id.*

⁶⁰ *Id.*

⁶¹ See MORRIS L. COHEN ET AL., *HOW TO FIND THE LAW* 84 (9th ed. 1989).

⁶² BARKAN ET AL., *supra* note 58, at 78 (emphasis removed).

⁶³ COHEN ET AL., *supra* note 61, at 84.

⁶⁴ BARKAN ET AL., *supra* note 58, at 78.

⁶⁵ *Id.*

⁶⁶ *Id.*

⁶⁷ *Id.* at 79.

⁶⁸ See Daniel Dabney, *The Universe of Thinkable Thoughts: Literary Warrant and West's Key Number System*, 99 *LAW LIBR. J.* 229, 229 (2007). West is now a unit of Thomson Reuters, but the name “West” is used in this article because that is the name most familiar to attorneys and best associated with the American Digest System and the Key Number System.

⁶⁹ THOMSON REUTERS, *WEST'S ANALYSIS OF AMERICAN LAW*, at V (2013) [hereinafter *WEST'S ANALYSIS OF AMERICAN LAW*].

contains over 400 topics, each of which is identified by its own topic number.⁷⁰ The topics are grouped into seven major categories—1. Persons, 2. Property, 3. Contracts, 4. Torts, 5. Crimes, 6. Remedies, and 7. Government.⁷¹ The topics are also divided into subtopics,⁷² which are organized in outline form.⁷³ This outline is known as the “analysis” of the topic.⁷⁴ Each subtopic is assigned a unique key number.⁷⁵ “A particular point of law, then, is known by its digest topic name [and number] and by its key number within that topic.”⁷⁶ For example, 331k18 is the key number that governs the avoidance of contractual releases on the basis of duress. Digest topic 331 is entitled “Release.”⁷⁷ And key number 18 within that topic is the subtopic “Duress.”⁷⁸

West’s system is intended to provide a key number “to every conceivable legal situation that could be treated in a case.”⁷⁹ As a result, there are over 100,000 individual key numbers.⁸⁰ Each headnote that West writes is classified using one or more key numbers⁸¹ and every case in West’s reporting system contains at least one headnote.⁸² In a typical year,

⁷⁰ *Id.* at XV–XVII.

⁷¹ *Id.* at V. A few of the topics are included in more than one category. For example, Insurance is listed under both Persons and Contracts. *Id.* at X–XI. Note also that all of the categories except for Torts and Crimes are broken into sub-categories. *Id.* at IX–XIV. To illustrate, there are four sub-categories under Contracts: (1) “Nature, Requisites, and Incidents of Agreements in General,” (2) “Particular Classes of Agreements,” (3) “Particular Classes of Implied or Constructive Contracts or Quasi Contracts,” and (4) “Particular Modes of Discharging Contracts.” *Id.* at XI.

⁷² *Id.* at V.

⁷³ BARKAN ET AL., *supra* note 58, at 83.

⁷⁴ *Id.*

⁷⁵ William H. Hilyerd, *Using the Law Library: A Guide for Educators Part V: Finding Legal Materials by Topic*, 34 J.L. & EDUC. 533, 545 (2005); Fritz Snyder, *The West Digest System: The Ninth Circuit and the Montana Supreme Court*, 60 MONT. L. REV. 541, 542 (1999).

⁷⁶ Snyder, *supra* note 75, at 542–43.

⁷⁷ WEST’S ANALYSIS OF AMERICAN LAW, *supra* note 69, at 1483.

⁷⁸ *Id.* at 1484. Note also that 331K18 is contained in the first section of the topic 331 outline—“Requisites and Validity [of Releases].”

⁷⁹ COHEN ET AL., *supra* note 61, at 84.

⁸⁰ WEST’S ANALYSIS OF AMERICAN LAW, *supra* note 69, at V.

⁸¹ COHEN ET AL., *supra* note 61, at 84, 86; Robert C. Berring, *Legal Research and Legal Concepts: Where Form Molds Substance*, 75 CALIF. L. REV. 15, 25 (1987); Snyder, *supra* note 75, at 541.

⁸² See COHEN ET AL., *supra* note 61, at 84; Snyder, *supra* note 75, at 543; *but cf.* Charles N.W. Keckler, *The Hazards of Precedent: A Parameterization of Legal Change*, 80 MISS. L.J. 105, 132 (2010) (“Many cases, even at the [state] Supreme Court level, have no Westlaw Key Number Treatment at all.”).

West writes 450,000 headnotes and adds more than 550,000 headnote classifications, with an average of eight headnote classifications per judicial opinion.⁸³

A case is added to the digest in the following manner. First, a West attorney-editor reads and analyzes the decision and crafts a headnote for each distinct point of law in the opinion.⁸⁴ Second, the case is sent to another West attorney-editor known as a “classifier” who assigns one or more key numbers to every headnote.⁸⁵ The classifiers “are experts in figuring out where points of law” fit in the Key Number System.⁸⁶ Third, and last, the classification decisions from the second stage are double-checked by more senior classification attorneys.⁸⁷ Note that all of the lawyers involved in this process receive considerable training in the performance of their tasks⁸⁸ and the average headnote classifier has more than twenty years of experience at West.⁸⁹

The West American Digest System was created in the late nineteenth century⁹⁰ and West began publishing the digest in 1887.⁹¹ The Key Number System was introduced two decades later in 1908.⁹² Since that time, the two

⁸³ Thomson Reuters, *The Westlaw Next Caselaw 22-Step Editorial Process 1*, available at http://info.legalsolutions.thomsonreuters.com/pdf/wln2/L-356219_v3.pdf. Some headnotes receive more than one key number classification. This explains why the number of classifications is greater than the number of headnotes added to the digest each year.

⁸⁴ BARKAN ET AL., *supra* note 58, at 79; Snyder, *supra* note 75, at 541.

⁸⁵ Dabney, *supra* note 68, at 229; Thomson Reuters, *West's Key Number System*, WESTLAW, <http://lawschool.westlaw.com/knumbers/> (last visited Mar. 22, 2014) [hereinafter *West's Key Number System*] (on file with author).

⁸⁶ *West's Key Number System*, *supra* note 85.

⁸⁷ Robert C. Berring, *Full-Text Databases and Legal Research: Backing Into the Future*, 1 HIGH TECH. L.J. 27, 32 (1986); E-mail from Erin Flood, Reference Attorney, Thomson Reuters, to author (Mar. 3, 2015, 18:28 CST) (on file with author).

⁸⁸ See *West's Key Number System*, *supra* note 85 (explaining that the headnote drafters are “highly trained”); *A Guide to Electronic Key Number Research on Westlaw.com*, WESTLAW, <https://lawschool.westlaw.com/marketing/display/RE/29> (last visited Feb. 20, 2015) (noting that key number classifiers are “trained specialists”).

⁸⁹ E-mail from Erin Flood, *supra* note 87.

⁹⁰ Susan Nevelow Mart, *The Case for Curation: The Relevance of Digest and Citorator Results in Westlaw and Lexis*, 32 LEGAL REF. SERV. Q. 13, 19 (2013).

⁹¹ Richard A. Danner, *Influences of the Digest Classification System: What Can We Know?*, 33 LEGAL REF. SERV. Q. 117, 122 (2014).

⁹² Douglas A. Hedlin, *A Citation History of Pine River State Bank v. Mettelle: A Study of Common Law Change, Judicial Influence, and the Birth of a Discipline*, 33 WM. MITCHELL L. REV. 297, 298 (2006) (citing WILLIAM W. MARVIN, WEST PUBLISHING COMPANY: ORIGIN, GROWTH, LEADERSHIP (1969)).

systems have been in a constant state of evolution, with West regularly adding, subtracting, and editing topics within the Key Number System to fit developments in the law and improve the general operation of the digest.⁹³ Indeed, West makes “hundreds or even thousands of changes” to the key number topics every year.⁹⁴

Sometimes a researcher can exploit data that a third party has already gathered and coded.⁹⁵ Using such data in empirical work is referred to as

⁹³ See BARKAN ET AL., *supra* note 58, at 88 (“West attempts to keep its key number system current with changes and developments in law by adding new topics and by expanding or reclassifying existing topics.”) (further describing these processes); Mart, *supra* note 90, at 19 (explaining that the system “has been evolving” since its creation); see also Dabney, *supra* note 68, at 240–42 (discussing an example of how a digest topic is modified); Snyder, *supra* note 75, at 543 (“The number of digest topics has remained fairly constant The number of key numbers within each topic, however, has increased significantly.”); *id.* at 558–61 (setting forth a history of digest topic changes).

⁹⁴ Thomson Reuters, *West Key Number System on Westlaw Next 1* (2012), available at <https://info.legalsolutions.thomsonreuters.com/pdf/wln2/L-374484.pdf>.

For most of the twentieth century, the West American Digest System was the principal tool for locating caselaw on a given subject. See COHEN ET AL., *supra* note 61, at 83. And it still plays a prominent, though reduced, role in legal research conducted by lawyers, judges, and law professors. See BARKAN ET AL., *supra* note 58, at 78 (explaining that computerized legal research has supplanted digests as the “most effective and efficient method for identifying relevant caselaw,” but noting the important role that digests still play); Jill L.K. Brooks, *Great Expectations: New Associates Research Skills from Law School to Law Firm*, 28 LEGAL REF. SERV. Q. 291, 295 (2009) (reporting the results of a survey of attorneys and law firm librarians; the respondents generally stated that new hires need to know how to use the Key Number System); Joseph A. Custer, *The Universe of Thinkable Thoughts versus the Facts of Empirical Research*, 102 LAW LIBR. J. 251, 259–60 (2010) (reporting the results of a survey of Kansas lawyers and law professors indicating that digests are still used in legal research by many individuals). The Key Number System thus remains an important device for gathering cases in traditional doctrinal scholarship. See, e.g., Jay M. Feinman, *Promissory Estoppel and Judicial Method*, 97 HARV. L. REV. 678, 679 n.5 (1984) (using the Key Number System to collect authorities for a doctrinal analysis of promissory estoppel).

For additional information on the American Digest System and the Key Number System, see WEST’S ANALYSIS OF AMERICAN LAW, *supra* note 69, at V–VII (explaining the organization of the Key Number System); Berring, *supra* note 87, at 29–38 (historical and operational overview of the West reporters and the Key Number System); Dabney, *supra* note 68 (assessing the ways in which the West American Digest System influences the law); Danner, *supra* note 91 (containing excellent overviews of (1) the history of the West Digest System, (2) the historical background of the legal classification movement of the late 19th and early 20th centuries, and (3) the literature regarding the influence of the West American Digest System on both the practice of law and the development of the law); Snyder, *supra* note 75 (discussing numerous historical and operational features of the West American Digest System); see also BARKAN ET AL., *supra* note 58, at 78–103 (general overview of digests and their use in legal research); COHEN ET AL., *supra* note 61, at 83–115 (same); Hilyerd, *supra* note 75, at 543–53 (same).

⁹⁵ LAWLESS ET AL., *supra* note 19, at 128.

“secondary data-analysis.”⁹⁶ This type of analysis saves time and money because “the hard work of recording the data has already been done.”⁹⁷ The West Key Number System is essentially a scheme for coding American caselaw.⁹⁸ That means that the American Digest is an immense database ripe for mining through secondary data analysis.⁹⁹ By crafting Westlaw search queries,¹⁰⁰ a researcher can accumulate a massive array of data coded by trained professionals using a system honed through more than a century of development.

B. Literature Review of Empirical Research Employing the Key Number System

Many legal scholars have used the American Digest System for empirical studies.¹⁰¹ But in most such work, key numbers are employed

⁹⁶ *Id.* at 129 (emphasis removed).

⁹⁷ *Id.* at 130; see generally *id.* at 128–34 (discussing research based on secondary data-analysis).

⁹⁸ See Ian Gallacher, *Mapping the Social Life of the Law: An Alternative Approach to Legal Research*, 36 INT’L J. LEGAL INFO. 1, 17 (2008) (explaining that headnotes “are coded according to the master list” of digest topics and key numbers) (emphasis added).

⁹⁹ See Daniel E. Ho & Kevin M. Quinn, *How Not to Lie with Judicial Votes: Misconceptions, Measurement, and Models*, 98 CALIF. L. REV. 813, 858 (2010) (explaining that the Key Number System might enable researchers to compile “a database of jurisprudentially meaningful information” without going through the “intensive and time-consuming” process of “reading each of . . . [the] cases to classify, disaggregate, and code the issues” discussed therein); Shapiro, *supra* note 3, at 528 n.213 (“One potential approach worth further investigation would be to turn to the West Key Number System [as a way of coding caselaw.]”); see also Hall & Wright, *supra* note 3, at 68 n.13 (“. . . 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.”); cf. Geis, *supra* note 4, at 477 (“While Lexis, Westlaw, and other avenues for retrieving the full text of judicial opinions have been available to legal scholars for decades, we have yet to fully exploit their empirical potential.”). Note that Lexis has a competing digest product that might also be useful for empirical research. See generally Mart, *supra* note 90, at 14, 18–21. However, I have not investigated this possibility.

¹⁰⁰ See THOMSON REUTERS WESTLAW, <https://next.westlaw.com>.

¹⁰¹ See, e.g., Richard Fossey, *Are Bankruptcy Courts Creating “the Certainty of Hopefulness” for Student Loan Debtors: Examining the “Undue Hardship” Rule*, in CONDEMNING STUDENTS TO DEBT: COLLEGE LOANS AND PUBLIC POLICY 161, 162, 171–72 (Richard Fossey & Mark Bateman eds., 1998); Abbe R. Gluck, *Intersystemic Statutory Interpretation: Methodology as “Law” and the Erie Doctrine*, 120 YALE L.J. 1898, 1925 & n.84 (2011); Dr. Joyce Holley & Danyne Holley, *Auditor Common Law Liability in the State Courts: A Recent (1980–94) Outcome Restatement and Perspectives of the Accounting and Legal Professions*, 6 U. MIAMI BUS. L. REV. 1, 2 & n.4, 3–5 (1997); Michael P. O’Shea, *The Steepness of the Slippery Slope: Second Amendment Litigation in the Lower Federal Courts and What It Has to Do With Background Recordkeeping Legislation*, 46 CONN. L. REV. 1381, 1383, 1415 (2014); David W. Romero & Francine Sanders Romero, *Precedent, Parity, and Racial Discrimination:*

only for data *collection*. After the opinions are gathered, they are read and coded by the authors or their research assistants.¹⁰² Using the Key Number System in this way saves time and money, to be sure. But in many contexts, the system can do more: It can be employed for both the collecting *and coding* of data. This dramatically increases the resources saved because the scholar need not manually code the cases to transform them into a standardized dataset. The digest classifications for each headnote have accomplished that already.

Law professors have produced a handful of studies in which they used the Key Number System for both data collection and coding. For example, to measure the role of pragmatic reasoning in federal circuit opinions concerning statutory interpretation, Frank Cross ran Westlaw searches to determine how often such opinions were coded with key number 361k181(2).¹⁰³ That key number fell under the topic “Statutes” and “roughly corresponds to the use of the absurdity doctrine in statutory interpretation,” a form of pragmatic reasoning.¹⁰⁴ Cross found that the number of opinions tagged with 361k181(2) went from less than five per year up to roughly twenty per year during the period of the study.¹⁰⁵ He concluded that these “data show a dramatic increase in references to pragmatism within circuit court opinions.”¹⁰⁶ Likewise, William Bratton and Michael Wachter used a Westlaw search for key numbers 101k1292 and 101k1475, which fall under the topic “Corporations and Business Organizations” and concern preferred stock, to support the conclusion that “Delaware courts have emerged as the dominant arbiters of preferred stock

A Federal/State Comparison of the Impact of Brown v. Board of Education, 37 LAW & SOC'Y REV. 809, 814 & n.5 (2003); Jane S. Schacter, *The Pursuit of “Popular Intent”: Interpretive Dilemmas in Direct Democracy*, 105 YALE L.J. 107, 110–11, 114–15 & n.27 (1995); Robert B. Thompson, *Piercing the Corporate Veil: An Empirical Study*, 76 CORNELL L. REV. 1036, 1036 n.1, 1044–45 (1991); Richard B. Wright, *Negligence in the Courts: Introduction and Commentary*, 77 CHI.-KENT L. REV. 425, 451–53 (2002).

¹⁰² Fossey, *supra* note 101; Gluck, *supra* note 101; Holley & Holley, *supra* note 101; O’Shea, *supra* note 101; Romero & Romero, *supra* note 101; Schacter, *supra* note 101; Thompson, *supra* note 101; Wright, *supra* note 101.

¹⁰³ Frank B. Cross, *What Do Judges Want?*, 87 TEX. L. REV. 183, 200–01 (2008).

¹⁰⁴ *Id.* at 200. West reorganized the “Statutes” topic after Professor Cross published his article and 361k181(2) was split up into multiple new key numbers. The absurdity doctrine is now covered by 361k1404. See WEST’S ANALYSIS OF AMERICAN LAW, *supra* note 69, at 1597.

¹⁰⁵ Cross, *supra* note 103, at 201.

¹⁰⁶ *Id.*

disputes.”¹⁰⁷ They found that “[s]ince 1980, 60% of cases keyed by West” with those two numbers were decided in Delaware; “New York came in a distant second with 20%.”¹⁰⁸

Professors of education have also conducted a number of studies using the Key Number System to measure trends in school-related litigation. To illustrate, Perry Zirkel and Sharon Richardson divided the key numbers under the topic “Schools” into five categories—(1) desegregation, (2) employees, (3) special education, (4) “Other Macro-Level,” and (5) “Other Students”—and ran separate Westlaw searches for cases classified with the numbers in each category.¹⁰⁹ They searched for both state and federal cases containing the relevant key numbers over five ten-year periods.¹¹⁰ Based on their search results, Zirkel and Richardson concluded that education litigation exploded in the 1960s and 1970s, but leveled off in the 1980s, especially in federal court, where the volume of litigation actually declined.¹¹¹ They also found that the leveling off or downturn in litigation applied to each of their categories except for special education.¹¹²

The other studies published by legal scholars¹¹³ and education professors¹¹⁴ have the same basic structure as the research conducted by

¹⁰⁷ William W. Bratton & Michael L. Wachter, *A Theory of Preferred Stock*, 161 U. PA. L. REV. 1815, 1821 & n.14 (2013).

¹⁰⁸ *Id.* at 1821 n.14.

¹⁰⁹ Perry A. Zirkel & Sharon N. Richardson, *The ‘Explosion’ in Education Litigation*, 53 WEST’S EDUC. L. REP. 767, 777 n.20 (1989). Note that since Zirkel and Richardson conducted the study, the “Schools” topic was eliminated and the subjects covered by that topic were moved to the topic “Education.” See WEST’S ANALYSIS OF AMERICAN LAW, *supra* note 69, at 619–37.

¹¹⁰ Zirkel & Richardson, *supra* note 109, at 777–78.

¹¹¹ *Id.* at 778–81, 789.

¹¹² *Id.*

¹¹³ See, e.g., George S. Geis, *Broadcast Contracting*, 106 NW. U. L. REV. 1153, 1155 & n.9 (2012) (supporting the proposition that claims of third-party beneficiary status are “attempted in numerous cases each year” via a Westlaw search for key number 95k187, “Agreement for benefit of third person,” which falls under the topic “Contracts”; the search identified more than one thousand judicial opinions containing that key number during a ten-year period); Keckler, *supra* note 82, at 108–09, 131–32 (In this study of the doctrine of stare decisis in Illinois, the author presented data regarding how often Illinois courts overrule precedents in different fields of law. To classify cases by field, the author relied upon the key numbers in the headnotes of each opinion.); David P. Leonard, *Character and Motive in Evidence Law*, 34 LOY. L.A. L. REV. 439, 441 & nn.7, 8 (2001) (using data from key number searches to support the contentions that Federal Rule of Evidence 404(b) “has been the subject of more appeals than any other evidence rule” and that a significant portion of those appeals concern motive); Gregory C. Lisby, *No Place in the Law: The Ignominy of Criminal Libel in American*

Cross, Bratton, Wachter, Zirkel, and Richardson. In each paper, the author(s) determined how many reported decisions were tagged with the relevant key numbers in the jurisdictions and time-period of the study. Then, based on the raw totals, the author(s) reached conclusions about trends in litigation, the frequency of different types of disputes, or the behavior of courts.

Jurisprudence, 9 COMM. L. POL'Y 433, 466–67 & nn.252–54 (2004) (manually counting decisions included in the American Digest under the key numbers relevant to criminal libel; concluding based on the data gathered that criminal libel prosecutions have been declining since the early twentieth century); Amy E. Sloan, *If You Can't Beat 'Em, Join 'Em: A Pragmatic Approach to Nonprecedential Opinions in the Federal Appellate Courts*, 86 NEB. L. REV. 895, 924–25 & n.203 (2008) (comparing a Westlaw search, which found that only 632 cases decided from 1941 through 2007 contained the key number concerning the authoritative value of prior opinions, to Shepard's Federal Citations, which at the time of the study contained 820 pages of entries identifying citations to nonprecedential circuit court opinions "published" in West's Federal Appendix; from this comparison, the author concluded that courts "frequently cite nonprecedential opinions without any discussion of their weight"); Snyder, *supra* note 75, at 547, 576–77 (presenting descriptive statistics regarding how often each digest topic appears in Montana Supreme Court and Ninth Circuit Court of Appeals cases; the author reviewed every decision by those two courts from 1945 through 1997; contending that this data is relevant to the question of which areas of law are most litigated, and thus possibly the most important, and that the data helps to explain "why certain courses are either required in law school or considered to be very important electives . . ."); see also Ho & Quinn, *supra* note 99, at 859–60, 875 (using the Key Number System's coding as part of an empirical study of the United States Supreme Court's statutory interpretation jurisprudence); William A. Wines, *Automobile Insurance Policies Build "Write-Away" Around Frolic and Detour, A Persistent Problem on the Highway of Torts*, 19 CAMPBELL L. REV. 85, 86–87 (1996) (using total pages in the West American Digest, rather than Westlaw or manual searches for cases, as a measure of the rate of litigation regarding the "frolic and detour" doctrine in tort law). Another interesting approach was used by Robert J. Levy in *Custody Investigations as Evidence in Divorce Cases*, 21 FAM. L.Q. 149 (1987). To support the proposition that "appellate cases from all jurisdictions raising custody investigations have grown significantly," Levy presented statistics indicating that the phrase "custody investigation" was mentioned in 960 West headnotes from 1951 to 1960, in 3,464 headnotes from 1961 through 1970, and in 8,227 headnotes from 1970–1980. *Id.* at 151 & n.11. Like the others authors of articles cited in this footnote, Levy relied upon coding performed by West. But instead of using the key number tagging completed by West's *classifiers* (the attorneys who assign key numbers to headnotes), he focused on the work performed by the attorney-editors who *write the actual headnotes*, treating their use of the phrase "custody investigation" as a form of coding of the opinions.

¹¹⁴ See, e.g., Clifford P. Hooker, *Teachers and the Courts, 1965–1986*, 48 WEST'S EDUC. L. REP. 7, 7–11 (1989); Michael Imber & David E. Gayler, *A Statistical Analysis of Trends in Education-Related Litigation Since 1960*, 24 EDUC. ADMIN. Q. 55, 58–59 (1988); Perry A. Zirkel & Brent L. Johnson, *The 'Explosion' In Education Litigation: An Updated Analysis*, 265 WEST'S EDUC. L. REP. 1, 1–6 (2011); *id.* at 1 n.2 (collecting additional authorities); Perry A. Zirkel, *The Volume of Higher Education Litigation: An Update*, 126 WEST'S EDUC. L. REP. 21, 21–27 (1998); Perry A. Zirkel, *Higher Education Litigation: An Overview*, 56 WEST'S EDUC. L. REP. 705, 705–08 (1990); Zirkel & Richardson, *supra* note 109, at 769–77 (reviewing prior research on education litigation, including several studies that used the West American Digest System).

My research uncovered only one scholar who has gone beyond reliance on raw totals and applied statistical analysis to the results of key number searches. Nicholas Georgakopoulos followed such a protocol in two studies of the caselaw on piercing the corporate veil. The first was part of an article arguing that veil piercing plays a critical role in contract disputes and is thus a desirable remedy in that type of lawsuit.¹¹⁵ Georgakopoulos began by reviewing the academic consensus that veil piercing is more justified when the plaintiff's claim is grounded in tort than when it is grounded in contract.¹¹⁶ He then presented his empirical study which was designed to test whether "courts and litigants demonstrate a bias in favor of piercing in contract disputes compared to tort disputes."¹¹⁷ His null hypothesis was that "litigants and courts have the same attitudes about piercing in contract as in tort."¹¹⁸ The study involved comparing (1) the ratio of all contracts cases to all tort cases, with (2) the ratio of veil piercing cases involving contract claims to veil piercing cases involving tort claims.¹¹⁹ "If no unique biases applied to piercing opinions, then one would expect that ratio to be constant. This would mean that whatever fraction of tort opinions contract opinions are, opinions about piercing in contract would be the same fraction of opinions about piercing in tort."¹²⁰ The ratio of contract decisions to tort decisions is the result of multiple dynamics. "If the same dynamics applied when the issue of piercing arose, then the same ratio would apply in the smaller set of piercing opinions."¹²¹

To measure the total number of contract cases, Georgakopoulos ran searches for all state and federal cases classified with a key number from topic 95, which is entitled "Contracts," in every year from 1947 to 2003.¹²² He followed the same protocol to determine the total number of torts cases, except he searched for opinions tagged with torts key numbers rather than

¹¹⁵ Nicholas L. Georgakopoulos, *Contract-Centered Veil Piercing*, 13 STAN. J.L. BUS. & FIN. 121, 123 (2007).

¹¹⁶ *Id.* at 124–27.

¹¹⁷ *Id.* at 127–30.

¹¹⁸ *Id.* at 127.

¹¹⁹ *Id.* at 127–30.

¹²⁰ *Id.* at 127.

¹²¹ *Id.*

¹²² *Id.* at 127–28 & n.22.

contracts key numbers.¹²³ He then ran Boolean term searches for decisions within this universe of contract and tort cases to determine how many referenced piercing the corporate veil.¹²⁴ I have set forth the raw totals from Georgakopoulos's study in Table 1.¹²⁵

Table 1. Number of opinions: contract, tort, and piercing

	No Piercing Reference	Piercing Reference	Predicted
Contract	61,040	475	290.53
Contract & Tort	4,375	68	20.98
Tort	109,901	285	520.40
All	175,316	828	

The ratio of opinions referring to piercing (828) to those not referring to piercing (175,316) is 0.47%.¹²⁶ To get the numbers in the third column ("Predicted"), multiply 0.47% by the numbers in the first column ("No Piercing Reference"). The predicted count is "the number of opinions that would mention piercing if they followed the allocation of the opinions that do not mention piercing."¹²⁷ As the raw totals make clear, the ratio of all contract cases to all tort cases is considerably smaller than the ratio of veil piercing cases involving contract claims to veil piercing cases involving tort claims. Georgakopoulos then ran a statistical test (the chi-test) to "calculate the probability that the observed allocation of *piercing* opinions can arise if the underlying generating dynamics or mechanism was the same" as for tort and contract cases *generally*.¹²⁸ This test resulted in a probability of less than one in a million or 0.00001%.¹²⁹ Georgakopoulos elaborates:

Statisticians can say this evidence rejects the null hypothesis with 99.9999% confidence. In lay terms, it is extremely improbable that so many piercing opinions are about contract disputes if piercing opinions are produced by the

¹²³ *Id.* Georgakopoulos used topics 272 ("Negligence"), 379 ("Torts") and selected portions of topic 48 ("Automobiles") to identify tort cases. *Id.* at 128 n.22.

¹²⁴ *Id.* at 128 & n.23. The search terms were "pierc! the corporate veil" and "disregard the corporate entity." *Id.* at 128 n.23.

¹²⁵ *Id.* at 129.

¹²⁶ *Id.*

¹²⁷ *Id.* at 128.

¹²⁸ *Id.* at 128–29 (emphasis added).

¹²⁹ *Id.* at 129.

same dynamics as overall contract and tort opinions. Piercing is invoked more in opinions about contract disputes than it is in opinions about tort.¹³⁰

Therefore, courts and litigants actually “demonstrate a bias in favor of piercing in contract disputes compared to tort disputes,”¹³¹ contrary to the academic consensus. In the rest of his article, Georgakopoulos set forth various theoretical arguments designed to explain why veil piercing is justified when parties attempt to enforce contract rights against corporations.¹³²

Georgakopoulos’s second study followed the same approach as the first, updating the research to cover the time period from 1947 through 2010, and reached essentially identical results.¹³³ This means that “either a different mechanism produces piercing disputes” or parties choose to pursue piercing more intensely in the contract context.¹³⁴ The balance of Georgakopoulos’s paper analyzed these two possibilities.¹³⁵

C. Methodological Issues

The scholarship discussed in the prior subpart demonstrates how some researchers have employed the Key Number System to collect and code data, sparing the authors considerable resources. In those studies, there was no need to analyze and code the relevant judicial opinions because West’s attorney-editors had already completed all or most of the required coding. Georgakopoulos’s two studies, in particular, also illustrate that using key number coding enables authors to analyze much larger datasets than in many other types of empirical research: The sample in his first study contained over 186,000 cases.¹³⁶ Most of the non-key number studies discussed above used samples in the hundreds.¹³⁷

¹³⁰ *Id.*

¹³¹ *Id.* at 127 (emphasis added).

¹³² *Id.* at 130–47.

¹³³ Nicholas L. Georgakopoulos, *Bankruptcy Veil-Piercing*, 27 EMORY BANKR. DEV. J. 471, 473–75 (2011).

¹³⁴ *Id.* at 475.

¹³⁵ *Id.* at 475–87. While conducting his analysis, Georgakopoulos included another small study using the key number system for data coding purposes. *See id.* at 480 n.24.

¹³⁶ Georgakopoulos, *supra* note 115, at 129.

¹³⁷ My study, while not as large as Georgakopoulos’s, involved a total of 8,113 cases. *See infra* Part V.B., Table 4; *see also* Hall & Wright, *supra* note 3, at 102–03 (empirical study of academic

However, employing the Key Number System in empirical research raises a variety of methodological issues beyond resource allocation. Any assessment of the system's value as a tool for data collection and coding must take into account these additional factors. Accordingly, this subpart discusses the other benefits and several limitations of using key numbers for empirical work.

There are multiple advantages to employing the Key Number System as a collection and coding device apart from the time and money that the system saves. First, key number classifications are assigned by parties who are "independent of the researcher and of any research hypothesis."¹³⁸ This eliminates author bias from the coding process and distinguishes coding performed by West's attorney-editors from coding completed by scholars or their research assistants.¹³⁹ Of course, in many circumstances, an author can use research assistants in ways that preserve the integrity of coding work. But that requires diligent attention to proper protocols.¹⁴⁰

Second, as noted above, West's attorney-editors have considerable training and experience in classifying headnotes with key numbers.¹⁴¹ Therefore, they are probably more accurate and reliable coders of caselaw than law students, the typical research assistants employed by legal scholars for coding work.¹⁴² The attorney-editors may even be superior to academic authors.

studies engaged in the systematic content analysis of judicial opinions; finding that only eleven of 114 studies using universal sampling coded more than 1000 cases; also finding that twenty-one coded 500–1000, thirty-nine coded between 100 and 300, and twenty-six coded fewer than 100); *id.* at 72 (explaining that in the 134 studies that comprised the full data set, the median number of cases coded was 252).

¹³⁸ Keckler, *supra* note 82, at 121 n.38. Keckler is referring to the coding done as part of West's citation service, KeyCite. *Id.* at 119, 121 n.38. But the same reasoning applies to the Key Number System, and Keckler uses that system for data coding later in his article. *See id.* at 131–32.

¹³⁹ In their survey of empirical studies involving case coding, Hall and Wright found that "[o]nly 22% (29/134) of the projects we reviewed primarily used student coders; in the rest, the authors appeared to do their own coding." Hall & Wright, *supra* note 3, at 109.

¹⁴⁰ For an excellent discussion of such protocols, *see id.* at 109–10.

¹⁴¹ *See supra* text accompanying notes 88–89.

¹⁴² *See* Hall & Wright, *supra* note 3, at 111 (explaining that some scholars have observed that law students make "a surprising number of mistakes" when coding judicial decisions) ("Student coders may lack the level of expertise needed to code reliably the more complex or subtle, yet more meaningful, aspects of judicial opinions."); *cf.* Ho & Quinn, *supra* note 99, at 858 (concluding that the Key Number System "has a chief advantage over the Supreme Court Database: unlike the issue codings therein, which are coded without legal expertise and represent *public policy* issues, Key Numbers are assigned

Third, multiple West attorney-editors participate in the process of assigning key numbers to each headnote.¹⁴³ This complies with the best practice that all data be coded by more than one person,¹⁴⁴ further enhancing the accuracy of the information contained in the American Digest System relative to information generated by author or research assistant coding.

Fourth, empirical studies using the Key Number System are exceptionally easy for other scholars to analyze and replicate.¹⁴⁵ An author's methodology often can be explained in the space of two pages or less. And other researchers can reproduce the data collection and coding simply by running identical searches on Westlaw.¹⁴⁶

Fifth, because of the ease of replicability, designing and executing comparable studies takes minimal effort. To illustrate, a subsequent author can run the same searches in different caselaw databases or during different

by attorneys using expertise to classify propositions of law"). For a general discussion of the process of selecting and training independent coders for empirical studies of caselaw and the strengths and weaknesses of using such coders, see Hall & Wright, *supra*, at 109–12.

¹⁴³ See *supra* text accompanying notes 85–87.

¹⁴⁴ See *supra* text accompanying note 31; see also Hall & Wright, *supra* note 3, at 111 (explaining that when a scholar codes her own data, "it is especially important from a scientific perspective to conduct reliability tests, by recruiting a colleague with similar expertise to double code at a least a subset of cases"); *id.* at 111 n.196 ("In our literature review, this was done in only five of the ninety projects where authors did some or all of their own coding."). Note, however, that West's attorney-editors work together in deciding on key number classification. And some sources contend that the best scientific practice is for coders to work independently. See Hall & Wright, *supra*, at 109–10 ("In theory, the most scientifically rigorous method is for researchers to train others to do the coding and for coders to work completely independently once they are trained."); *id.* at 110 n.186 ("Experts caution against allowing coders to confer with each other once they are trained because this undermines the independence of their judgment, thus calling into question the objectivity and reproducibility of their coding."); *but see id.* at 116–17 (discussing various forms of coding where the coders work together as a group to make classification decisions).

¹⁴⁵ Cf. Keckler, *supra* note 82, at 121 n.38 (explaining that one of the virtues of using West's citation service, KeyCite, as a device for coding caselaw is that anything done with that system "is easily replicable"); see generally Hall & Wright, *supra* note 3, at 105–06 ("An essential attribute of scientific objectivity is *replicability*—the ability at least in theory, to test a research project by reproducing its findings using the same methods. Replicability is the overriding reason for using systematic content analysis. This is what confers scientific status on research findings."); Korobkin, *supra* note 4, at 1051 (explaining the importance of describing one's research methodology so that others can replicate the study).

¹⁴⁶ See also Hall & Wright, *supra* note 3, at 106 (explaining that many empirical studies of caselaw are not fully replicable because, after running electronic searches, the author uses additional subjective criteria to "screen out cases").

time periods. Alternatively, the scholar can modify the searches and run them in the same (or different) databases and time periods.

Sixth, when engaging in secondary data analysis, “researchers need to know as much as possible about the data,” including who collected it, why it was gathered, the procedures employed in the collection process, and the coding scheme that was used to standardize the information.¹⁴⁷ The American Digest System meets all of these requirements.

Between the resources saved and the six advantages discussed immediately above, it is clear that there are substantial benefits to using key numbers in empirical scholarship. But there are also critical limitations. The first and most important concerns the scope and content of the American Digest System. The cases that receive headnotes and key numbers are a tiny subset of the broader populations that researchers typically wish to study. And there is compelling evidence that the digested opinions are not a representative sample of the larger categories.

Starting with the quantitative aspect of this limitation, West normally writes headnotes and assigns key numbers only for “published” or “reported” decisions¹⁴⁸—i.e., cases selected for inclusion in its National Reporter System.¹⁴⁹ Most opinions printed in that system are issued by

¹⁴⁷ LAWLESS ET AL., *supra* note 19, at 131.

¹⁴⁸ I use the terms “published” and “reported” interchangeably in this article.

¹⁴⁹ See BARKAN ET AL., *supra* note 58, at 86 (“The *American Digest System* has as its core a massive set of materials in several units known as the *Decennial Digests* and their companion volumes containing later information, the *General Digest*. . . . Each volume [of the *General Digest*] consists of all the headnotes taken from all the units of the *National Reporter System* for the period covered.”) (emphasis in original); COHEN ET AL., *supra* note 61, at 86, 90 (“West . . . publishes two interlocking, comprehensive services: a system of case reporting which contains the full text of appellate . . . decisions from state and federal courts, and a digest structure providing for the classification . . . of the points of law determined in all judicial decisions reported by West.”); *id.* at 50 (noting that West adds headnotes and key numbers to all opinions included in the National Reporter System); see also Snyder, *supra* note 75, at 542 (“The American Digest System includes the headnotes from the *published* cases of all federal courts and the *published* cases of the appellate courts of every state.”) (emphasis added); Joseph L. Gerken, *A Librarian’s Guide to Unpublished Judicial Opinions*, 96 LAW LIBR. J. 475, 480 (2004) (noting that published decisions are the rulings “included in a West print reporter”). However, there are an increasing number of exceptions. “Selected state appellate and lower court unpublished decisions receive [headnote and key number] enhancements” when West, in its editorial judgment, determines that the opinion warrants inclusion in the digest. E-mail from Adrienne Stanley, Reference Attorney, Thomson Reuters, to Author (July 22, 2015, 9:17 CST) (on file with author). And during my research, I came across several opinions with headnotes and key numbers that were not printed in any volume of the National Reporter System. See, e.g., *Nat. Lumber Co. v. Brainerd*, 76 Mass. App. Ct. 1102, 2009 WL 4824781 (Mass. App. Ct. Dec. 16, 2009).

federal and state appellate courts.¹⁵⁰ Exceptionally few state trial court decisions are published.¹⁵¹ And while a significant number of federal trial court rulings are included in printed volumes like the Federal Supplement and the Federal Rules Decisions,¹⁵² they are still a small fraction of the orders issued by district court judges.¹⁵³ Very few lawsuits result in an appellate opinion.¹⁵⁴ Combining this point with the fact that most reported

¹⁵⁰ Lisa R. Pruitt, *Her Own Good Name: Two Centuries of Talk About Chastity*, 63 MD. L. REV. 401, 433 n.159 (2004) (observing that the “West reporting system contains mostly appellate level cases”); see Wayne Klein, *The Idaho Securities Act: An Analysis of Idaho Courts’ Securities Opinions*, 29 IDAHO L. REV. 95, 111 (1992/1993) (“Only appellate decisions are routinely reported.”); see also COHEN ET AL., *supra* note 61, at 83 (“The West Publishing Company’s National Reporter System . . . is a comprehensive scheme of publication of appellate court decisions from all jurisdictions.”).

¹⁵¹ Gerken, *supra* note 149, at 479 (observing that “very few [state] trial level decisions are published”); Klein, *supra* note 150, at 111 (“Most significantly, state trial decisions are not generally reported.”); Pruitt, *supra* note 150, at 433 n.159 (“[M]ost trial level cases are not reported.”). Note that some states do publish a significant number of trial court rulings. For example, opinions from trial courts in New York, New Jersey, and Pennsylvania are reported sufficiently often to warrant citation rules. See THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION 277, 280, 289 (Columbia Law Review Ass’n et al. eds., 20th ed. 2015). By comparison, California, Illinois, and Texas have no such rules. See *id.* at 252, 262, 295.

¹⁵² To illustrate, at the time I wrote this section of the article, West’s Federal Supplement 2d had reached volume 217. Note also that a significant number of opinions issued by district court judges and bankruptcy court judges are contained in West’s Bankruptcy Reports.

¹⁵³ See David A. Hoffman et al., *Docketology, District Courts, and Doctrine*, 85 WASH. U. L. REV. 681, 708–10 (2007) (study of 980 cases from four federal districts containing 1,091 substantive orders, such as summary judgment rulings, and 4,631 ministerial orders, such as scheduling orders; 178 orders, all of them substantive, were available on Westlaw; that constitutes 16% of the substantive orders and 3% of all orders; in addition, only 18% of the 980 cases resulted in an order contained on Westlaw; note further that the category of opinions available on Westlaw is broader than the category of opinions contained in a West reporter); Margo Schlanger & Denise Lieberman, *Using Court Records for Research, Teaching, and Policy Making: The Civil Rights Clearinghouse*, 75 UMKC L. REV. 155, 164–65 (2006) (finding that roughly 2.3% of district court rulings that disposed of a case in fiscal 2004 were published in the Federal Supplement or the Federal Rules Decisions, and 8.7% of such rulings were available on Westlaw; this result was derived by comparing (1) all 27,890 district court opinions posted on Westlaw during fiscal 2004 (note that Westlaw also identifies whether the decision was printed in a reporter), with (2) the Administrative Office of the U.S. Courts’ “reported number of civil and criminal cases ‘terminated’ in each district” for the same period); Peter Siegelman & John Donohue, *Studying the Iceberg from Its Tip: A Comparison of Published and Unpublished Employment Discrimination Cases*, 24 LAW & SOC’Y REV. 1133, 1138, 1141 (1990) (concluding, in this study of 4,310 employment civil rights cases filed in the Northern District of Illinois between July 1, 1972 and December 31, 1986, that only 20.1 percent, or 867 cases, generated a published opinion; the study defined published to mean available on Lexis, which is a broader category than all cases contained in a West reporter).

¹⁵⁴ George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. LEGAL STUD. 1, 2 (1984) (“It is well known, however, that only a very small fraction of disputes comes to trial and an even smaller fraction is appealed.”) (collecting authorities); Mark R. Brown & Andrew C. Greenberg, *On Formally Undecidable Propositions of Law: Legal Indeterminacy and the Implications*

cases are decided by courts of appeal establishes that published decisions containing headnotes and key numbers are a tiny subset of all actions brought before civil and criminal tribunals.

The set of cases added to the American Digest System is actually even smaller than I suggested in the previous paragraph. That is because most states have rules that limit the publication of opinions issued by intermediate courts of appeal.¹⁵⁵ As a result, the majority of rulings made by such courts not reported.¹⁵⁶ Consider Illinois: In 2013, only 531 out of 2248 cases resolved on the merits in that jurisdiction's intermediate appellate courts were disposed of by a published opinion. The rest were concluded via some other type of order.¹⁵⁷ For criminal matters, it was 314 reported decisions out of 2386 merits resolutions.¹⁵⁸

of Metamathematics, 43 HASTINGS L.J. 1439, 1453 (1992) (“Of those cases that do wind up in court, only a small portion are appealed and reported.”); Eli Wald & Russell G. Pearce, *Making Good Lawyers*, 9 U. ST. THOMAS L.J. 403, 417 (2011) (stating that only a “miniscule subset of legal matters . . . result in appellate decisions”).

¹⁵⁵ See, e.g., ILL. S. CT. R. 23(a) (providing that a case pending before an Illinois Appellate Court may only be disposed of via a publishable opinion if (1) “the decision establishes a new rule of law or modifies, explains or criticizes an existing rule of law; or (2) the decision resolves, creates, or avoids an apparent conflict of authority within the Appellate Court”); see generally Melissa M. Serfass & Jessie L. Cranford, *Federal and State Court Rules Governing Publication and Citation of Opinions: An Update*, 6 J. APP. PRAC. & PROCESS 349, 358–94 (2004) (setting forth a comprehensive list of such rules as of 2004). Unpublished appellate orders are generally not precedential and, in most states, litigants are forbidden from citing such rulings except in a narrow set of circumstances. See *id.*; e.g., ILL. S. CT. R. 23(e)(1) (permitting citation to unpublished decisions only “to support contentions of double jeopardy, *res judicata*, collateral estoppel or law of the case”). Note, however, that some states have eliminated the distinction between published and unpublished decisions at the appellate level. For example, Arkansas previously followed the approach used by other jurisdictions. See Jillian R. Jones, Comment, *Bound by Precedent: Arkansas Practitioners Win the Debate over Unpublished Decisions*, 63 ARK. L. REV. 619, 621 (2010) (“Opinions intended as precedent were designated as ‘published’ and appeared in the *Arkansas Reports*; however, opinions deemed as lacking precedential value were excluded from the *Arkansas Reports* and marked as ‘not designated for publication.’”). But in 2009, the Arkansas Supreme Court “adopted electronic publication and distribution for the official reports of all Arkansas appellate decisions . . . and abandoned its distinction between published and unpublished opinions and, most notably, declared that every Supreme Court and Court of Appeals opinion issued after July 1, 2009, had precedential value.” *Id.* at 619; see also RULES OF THE S. CT. AND CT. OF APPEALS OF THE STATE OF ARK. 5-2, available at <https://courts.arkansas.gov/rules-and-administrative-orders/rules-of-the-supreme-court-and-court-of-appeals-of-the-state-of-arkansas>.

¹⁵⁶ Pruitt, *supra* note 150, at 433 n.159.

¹⁵⁷ ADMINISTRATIVE OFFICE OF THE ILLINOIS COURTS, 2013 ANNUAL REPORT OF THE ILLINOIS COURTS 142, 144, available at http://www.illinoiscourts.gov/supremecourt/annualreport/2013/statsum/2013_statistical_summary.pdf. Note that the Illinois Appellate Court disposes of many cases on jurisdictional grounds, pursuant to a motion to dismiss by the appellant, or via a similar proceeding

Federal circuit courts also have local rules or practices that limit the issuance of published opinions,¹⁵⁹ most of which were initially adopted in the late 1970s.¹⁶⁰ And a substantial majority of the orders issued by these courts are unpublished.¹⁶¹ However, in 2001, West started producing a new volume called the Federal Appendix.¹⁶² That series contains “unpublished”¹⁶³ orders issued from 2001 forward by federal appeals courts and most of the decisions printed therein receive headnote and key number treatment.¹⁶⁴ Thus, since 2001, West has included a much higher portion of

that does not result in a merits ruling. If these orders are included in the totals, then 531 out of 4370 appellate cases were resolved via a published decision. *Id.*

¹⁵⁸ *Id.* Including the non-merits orders changes the totals to 314 out of 3384.

¹⁵⁹ See generally David R. Cleveland, *Local Rules in the Wake of Federal Rule of Appellate Procedure 32.1*, 11 J. APP. PRAC. PROCESS. 19, 61–72 (2010). However, while unpublished appellate rulings are generally non-precedential, *id.*, parties may now cite such opinions in the federal system, see FED R. APP. P. 32.1, contrary to the practice in state court, see *supra* note 155.

¹⁶⁰ See Gerken, *supra* note 149, at 477–78.

¹⁶¹ See Hon. Harry T. Edwards & Michael A. Livermore, *Pitfalls of Empirical Studies that Attempt to Understand the Factors Affecting Appellate Decisionmaking*, 58 DUKE L.J. 1895, 1923 (2009) (“[A] huge percentage of [federal] courts of appeals decisions are reported but unpublished. In fact, in 2007, less than 17 percent of all opinions in the [federal] courts of appeals were published.”); ADMIN. OFFICE OF THE U.S. COURTS, TABLE 2.5, U.S. COURTS OF APPEALS JUDICIAL FACTS AND FIGURES (Sept. 30, 2013), <http://www.uscourts.gov/statistics/table/25/judicial-facts-and-figures/2013/09/30> (last visited June 24, 2015) (setting forth data indicating that in 2011, 2012, and 2013, a total of 85.0%, 81.4%, and 82.3% of orders issued by the federal circuit courts of appeal were unpublished in cases terminated on the merits after an oral hearing or submission on briefs).

¹⁶² Laura K. Donohue, *The Shadow of State Secrets*, 159 U. PA. L. REV. 77, 81 n.8 (2010); Gerken, *supra* note 149, at 475.

¹⁶³ I placed the word “unpublished” in quotation marks because the rulings contained in the Federal Appendix are clearly “published” under the common understanding of that term. See also Perry A. Zirkel with Amanda C. Machin, *The Special Education Case Law “Iceberg”: An Initial Exploration of the Underside*, 41 J.L. & EDUC. 483, 484 (2012) (“In recent years, the concept of ‘published’ has become more flexible and expansive, ultimately affording a larger available corpus of judicial decisions for empirical analysis.”).

¹⁶⁴ See Thomson Reuters, *Federal Appendix (National Reporter System)*, <http://legal.solutions.thomsonreuters.com/law-products/Reporters/Federal-Appendix-National-Reporter-System/p/100000796> (last visited June 24, 2015) [hereinafter *Federal Appendix*] (“The *Federal Appendix* . . . covers opinions and decisions from 2001 to date issued by the U.S. courts of appeals that are not selected for publication in the *Federal Reporter*.”). In particular, the *Federal Appendix* “includes orders routinely made accessible as unpublished by federal appellate courts.” E-mail from Adrienne Stanley, *supra* note 149; accord Gerken, *supra* note 149, at 479 (“According to West, the *Federal Appendix* will include the full text of unpublished decisions to which the publisher has access. . . .”). Some types of orders, such as rulings addressing various motions, are typically inaccessible. These decisions are generally not printed in the *Federal Appendix*. See E-mail from Adrienne Stanley. Finally, all opinions “that have reasoning/legal conclusions received editorial enhancements.” *Id.*; see also *Federal Appendix*

federal appellate decisions in the American Digest System than other types of rulings.

To recap: (1) only a tiny percentage of trial court orders are reported; (2) most state appellate court rulings are unreported; (3) the bulk of federal appellate decisions were not reported from the late 1970s through 2001; and (4) a small fraction of lawsuits culminate in an appellate opinion. To these, we must add the following: (5) the vast majority of human behavior that is governed or influenced by legal norms does not result in litigation.¹⁶⁵ These five points together entail that the judicial opinions with key number coding are a microscopic subset of all lawsuits filed and orders issued, and an even tinier fraction of other potential categories of empirical study, such as disputes, injuries, accidents, crimes, contracts, or transactions.¹⁶⁶

The small number of cases contained in the digest would not be a significant issue if these decisions were a representative subdivision of the broader classifications that researchers are interested in analyzing.¹⁶⁷ But a considerable body of literature has found that reported opinions are not representative of any such larger group.

Beginning with the appellate level, state and local federal rules ordinarily permit intermediate courts of appeal to publish a decision only if

(“Cases published in this product series are enhanced with headnotes, key numbers, and synopses prepared by West’s internal staff of highly trained attorney editors.”).

¹⁶⁵ Hall & Wright, *supra* note 3, at 104 (“[M]ost human interactions do not produce disputes, only some disputes result in legal claims”); Lonny Sheinkopf Hoffman, *The Case Against Vicarious Jurisdiction*, 152 U. PA. L. REV. 1023, 1082 (2004) (noting that “the total number of lawsuits filed is . . . only a small percentage of the total quantum of behavior affected by prevailing legal norms”).

¹⁶⁶ See Brown & Greenberg, *supra* note 154, at 1453 (explaining that “reported cases represent only a very small percentage of the controversies and transactions that occur in the United States each day”); Hoffman, *supra* note 165, at 1082 (“[R]eported decision are but a small percentage of the total number of lawsuits filed, and the total number of lawsuits filed is, in turn, only a small percentage of the total quantum of behavior affected by prevailing legal norms.”); see also Siegelman & Donohue, *supra* note 153, at 1133 (“One of the few uncontested truths produced by the application of social science to law is that only a tiny subset of the ‘action’ in the legal system is revealed in appellate cases. Most potential disputes never get defined by the actors as such, most actual disputes don’t go to court, most court cases are settled rather than adjudicated, and most adjudicated cases are not appealed.”) (citing Richard E. Miller & Austin Sarat, *Grievances, Claims, and Disputes: Assessing the Adversary Culture*, 15 LAW & SOC’Y REV. 525, 544 (1981)).

¹⁶⁷ Cf. Schlanger & Lieberman, *supra* note 153, at 165 (“It would matter less that so few cases are observable from the judicial opinions if those cases fairly represented the entire docket.”).

it has significant precedential value.¹⁶⁸ These rules provide a firm basis for hypothesizing that reported appellate cases are different from those that are unreported. And there is much evidence supporting this proposition. For example, it is well known that “the judgments rendered in unpublished decisions are largely unanimous, and . . . typically involve more straightforward applications of law.”¹⁶⁹ In addition, studies analyzing appellate rulings have found that: (1) the frequency with which various types of parties appear and their rates of success differ between reported and unreported cases;¹⁷⁰ (2) reversal rates are higher in published decisions;¹⁷¹ and (3) panel composition influences whether an order is reported.¹⁷² Note also that the percentage of appellate rulings that are published varies by federal circuit,¹⁷³ as does the percentage of reversals that are published.¹⁷⁴ This means that a national sample of reported federal appellate cases will often be geographically biased. In sum, there is little question that appellate rulings with headnotes and key numbers are not representative of appellate matters as a whole.

¹⁶⁸ See Shenoa L. Payne, *The Ethical Conundrums of Unpublished Opinions*, 44 WILLAMETTE L. REV. 723, 726–27 (2008); Cleveland, *supra* note 159, at 61–72; Serfass & Cranford, *supra* note 155, 358–94; *see also, e.g.*, ILL. S. CT. R. 23(a), quoted in note 155 *supra*.

¹⁶⁹ Edwards & Livermore, *supra* note 161, at 1923.

¹⁷⁰ See Donald R. Songer & Reginald S. Sheehan, *Who Wins on Appeal? Upperdogs and Underdogs in the U.S. Courts of Appeals*, 36 AM. J. POL. SCI. 235, 249–50, 252–53 (1992).

¹⁷¹ *See id.* at 249, 257 (referring to an unpublished study conducted by one of the authors); Robert A. Mead, *Unpublished Opinions as the Bulk of the Iceberg: Publication Patterns in the Eighth and Tenth Circuits of the United States Courts of Appeals*, 93 LAW LIBR. J. 589, 603, 605 (2001) (study comparing published and unpublished decisions of the Eighth and Tenth Circuits for the first six months of 2000).

¹⁷² Carlos Berdejo, *It's the Journey, Not the Destination: Judicial Preferences and the Decision-Making Process*, 51 U. LOUISVILLE L. REV. 271, 274 n.13 (2013) (explaining that “whether a decision is published or not may be correlated with the composition of the panel writing the opinion” and collecting authorities that support this conclusion).

¹⁷³ ADMIN. OFFICE OF THE U.S. COURTS, TABLE 2.5, U.S. COURTS OF APPEALS JUDICIAL BUSINESS (Sept. 30, 2014), <http://www.uscourts.gov/statistics/table/b-12/judicial-business/2014/09/30> (last visited June 24, 2015) (setting forth data indicating that during the 12 months ending September 30, 2014, the percentage of opinions or orders that were unpublished for the twelve federal circuit courts of appeal ranged from 54.1% in the First Circuit to 93.8% in the Fourth Circuit).

¹⁷⁴ *See* Mead, *supra* note 171, at 603, 605 (in this study comparing published and unpublished decisions of the Eighth and Tenth Circuits for the first six months of 2000, the author presented data indicating that the Eighth Circuit published a considerably higher percentage of its reversals than did the Tenth Circuit).

Now consider trial courts. Federal district court judges possess nearly “complete discretion regarding whether to publish, or even write, an opinion in a particular case.”¹⁷⁵ And they do not randomly select rulings for the reporters.¹⁷⁶ Instead, they are influenced by “formal rules, court culture, personal predilections, or strategic considerations.”¹⁷⁷ It should thus not be surprising that multiple empirical studies of the federal trial courts have found that reported cases differ from unreported matters.¹⁷⁸ To illustrate, Peter Siegelman & John Donohue compared a random sample of employment discrimination cases that involved at least one published opinion to a random sample of such cases that did not produce any published decision.¹⁷⁹ They found numerous differences between the two sets of lawsuits.¹⁸⁰ For example, the court files for matters with a published opinion were two and one-half times as thick.¹⁸¹ The proportion of suits that involved a certified class action was six times higher for the reported set (13.1% versus 1.9%).¹⁸² And both the total dollars awarded (\$606,424 versus \$12,545) and the dollars awarded per plaintiff (\$49,907 versus \$12,375) were much higher in the published cases.¹⁸³ Siegelman & Donohue’s results also indicated that lawsuits with a published opinion tend to last longer, “to include a different mix of plaintiff occupations, to proceed at a different pace through the legal system, and to end in different kinds of outcomes.”¹⁸⁴

¹⁷⁵ Ahmed E. Taha, *Data and Selection Bias: A Case Study*, 75 UMKC L. REV. 171, 174 (2006).

¹⁷⁶ Pauline T. Kim et al., *How Should We Study District Judge Decision-Making?*, 29 J.L. & POL’Y 83, 97 (2009).

¹⁷⁷ *Id.*

¹⁷⁸ For an excellent yet brief survey of the literature, see Schlanger & Lieberman, *supra* note 153, at 165–67 (among other things, the authors specifically note that “there is now voluminous evidence that [judges] choose to devote the time to fully developed opinion writing in non-representative ways”).

¹⁷⁹ Siegelman & Donohue, *supra* note 153, at 1150. The two samples contained 114 and 105 cases, respectively. *Id.* And all of the cases were filed in the Northern District of Illinois. *Id.* at 1138. Note that the authors defined “published” to mean available on Lexis. *Id.* at 1138. This definition includes cases beyond those contained in West reporters.

¹⁸⁰ *Id.* at 1150–56.

¹⁸¹ *Id.* at 1150.

¹⁸² *Id.*

¹⁸³ *Id.* at 1151–52.

¹⁸⁴ *Id.* at 1156. In the same article, the authors conducted an additional study finding that district courts vary in the percentage of orders they publish. *See id.* at 1142–44. Accordingly, relying “solely on cases with published opinions will generally produce a geographically skewed sample of all cases filed” in district courts. *Id.* at 1144. Multiple other studies have reached the same conclusion. *See Hoffman et*

In another study, Susan Olson found critical variation in the types of federal cases that produce a published opinion.¹⁸⁵ Of the matters in her dataset,¹⁸⁶ very few real property (0%), personal property (0%), personal injury (0.7%), and labor disputes (1.9%) resulted in a published decision, whereas bankruptcy (11.8%), civil rights (12.2%), and federal regulatory cases (17.7%) produced a significant number.¹⁸⁷ Olson's findings suggest that samples of published decisions will often contain subject-matter biases.¹⁸⁸

Further illumination regarding the unrepresentative nature of reported decisions can be found in the dispute resolution literature. Researchers in this field often conceptualize their various categories of study as constituting a pyramid. The pyramid is best understood as consisting of the following five levels (proceeding from top to bottom): (1) appellate cases,

al., *supra* note 153, at 698 (collecting authorities); *id.* at 710 (setting forth the authors' own research); Schlanger & Lieberman, *supra* note 153, at 166–67 (discussing the authors' own study).

¹⁸⁵ Susan M. Olson, *Studying Federal District Courts Through Published Cases: A Research Note*, 15 JUST. SYS. J. 782 (1992). Like Siegelman and Donohue, Olson defined "published" to mean accessible on Lexis. *Id.* at 789.

¹⁸⁶ The data for this study came from "a systematic sample of every fourth civil case terminated in Minnesota federal district court from July 1, 1982 through June 30, 1984." *Id.* at 788. This resulted in a dataset of 697 cases. *Id.* at 790.

¹⁸⁷ *Id.* at 790, 791.

¹⁸⁸ *Accord* Hoffman et al., *supra* note 153, at 719 (finding, in this study of 980 cases from four federal districts, that contracts and labor lawsuits produced above-average numbers of opinions available on Westlaw and torts and habeas prisoner actions produced below-average numbers of such opinions; all four findings were statistically significant); *see also* Schlanger & Lieberman, *supra* note 153, at 166 ("Besides outcome skews, it is plausible that [published] opinions are about different kinds of cases than are typical on the docket—I would hypothesize that they are more likely to concern complex than simple cases, more likely to be issued in large stakes than small stakes cases, and so on."). For other studies concluding that reported and unreported cases in federal district court differ, *see* C.K. ROWLAND & ROBERT A. CARP, *POLITICS AND JUDGMENT IN FEDERAL DISTRICT COURTS* 123–28 (1996) (finding that the influence of political ideology appears to be greater in published district court opinions than in unpublished ones); Evan J. Ringquist & Craig E. Emmert, *Judicial Policymaking in Published and Unpublished Decisions: The Case of Environmental Civil Litigation*, 52 POL. RES. Q. 7, 7, 16–17, 27–33 (1992) (In this study of 1074 civil enforcement lawsuits brought by the Environmental Protection Agency, the authors found (1) that significantly higher penalties were awarded in published cases (\$62,380) than in all actions (\$28,294), and (2) that political considerations, such as the party of the appointing President, have greater influence on the decision in published cases); *see also* Schlanger & Lieberman, *supra*, at 165 (discussing four reasons why published trial court rulings are not representative of all trial court matters); *but see* Zirkel & Machin, *supra* note 163, at 499–507 (finding that unpublished decisions "are not dramatically different from published decisions" based on a sample of sixty-four cases (forty-three published and sixteen unpublished) brought under the Individuals with Disabilities in Education Act in federal district court during the year 2000).

(2) litigated cases that reach judgment before a trial court, (3) filed cases, (4) disputes, and (5) human interactions.¹⁸⁹ Each level in the pyramid is much smaller than the one immediately below it.¹⁹⁰ For example, since few trial court judgments are appealed, appellate cases constitute a small subset of litigated cases.¹⁹¹ Likewise, most disputes do not result in the filing of a lawsuit.¹⁹²

Critically, the consensus among dispute resolution scholars is that each level in the pyramid is not a representative subset of the group beneath it. To illustrate, it is widely accepted that appealed cases are different from actions litigated to judgment before a lower court.¹⁹³ Similarly, matters

¹⁸⁹ See, e.g., Taha, *supra* note 175, at 171–72 & n.1. The pyramid is sometimes called the “grievance pyramid” and it can be conceptualized in a variety of ways, adding considerably more layers. See, e.g., Kevin M. Clermont & Theodore Eisenberg, *Litigation Realities*, 88 CORNELL L. REV. 119, 135–36 (2002).

¹⁹⁰ Clermont & Eisenberg, *supra* note 189, at 135–36; Siegelman & Donohue, *supra* note 153, at 1133 (the relevant language from page 1133 of this article is quoted in footnote 166, *supra*); Taha, *supra* note 175, at 171–73.

¹⁹¹ Taha, *supra* note 175, at 173.

¹⁹² *Id.* at 172 & n.2 (citing David M. Trubek et al., *The Costs of Ordinary Litigation*, 31 UCLA L. REV. 72, 85–85 (1983) (survey of 5,000 households finding that only 11.2% of disputes resulted in a lawsuit)).

¹⁹³ Siegelman & Donohue, *supra* note 153, at 1134 (“[S]ocial scientists studying the law have come to appreciate that the process by which parties select disputes for litigation or appeal will not, in general, produce a random (representative) sample of all disputes or cases. . . . The unrepresentative nature of appellate cases is by now widely accepted among social scientists.”); Christina L. Boyd & James F. Spriggs II, *An Examination of Strategic Anticipation of Appellate Court Preferences by Federal District Court Judges*, 29 WASH. U. J.L. & POL’Y 37, 47 (2009) (“After all, cases that are appealed and advance into each higher tier represent a non-random sample of all cases that are litigated.”); Richard S. Grunner, *How High is Too High?: Reflections on the Sources and Meaning of Claim Construction Reversal Rates at the Federal Circuit*, 43 LOY. L.A. L. REV. 981, 1012 (2010) (“[A]ppealed cases involve a nonrandom subset of tried cases, most often involving tried cases that still have material uncertainties.”).

One older study of eleven states over a ten-year period (ending in 1978) found a strong correlation between the total level of federal and state trial-court cases and the total level of appellate cases. See Imber & Gayler, *supra* note 114, at 57–58. Given the strength of the correlation, “and in the absence of any empirical or theoretical reason to believe otherwise,” the authors concluded that “it is reasonable to accept the hypothesis that trends in appellate court cases in a particular category are a strong indicator of trends in trial court caseloads in the same category.” *Id.* at 58. The authors thus decided that it was appropriate to use the results of key number searches on Westlaw as a proxy for total levels of education-related litigation. *Id.* at 57–59. Unfortunately, there are now strong “empirical and theoretical reasons” to believe that various categories of cases result in reported rulings at different rates. For example, I discussed above two studies that found that different types of lawsuits produce published opinions at differing frequencies. See *supra* notes 185–88 and accompanying text. And both of the articles containing these studies explored various theories and supporting empirical evidence that could

litigated to judgment are not reflective of all filed cases.¹⁹⁴ One key reason is that lawsuits that reach judgment via dispositive motion or trial are unlike those that settle.¹⁹⁵

If we combine the studies of reported cases discussed above with the insights of the dispute resolution literature, we see that the problem of unrepresentative samples is compounded. Published rulings are not representative of their associated level in the pyramid. And each level in the pyramid is not representative of the levels beneath it.¹⁹⁶

In empirical work, “selection bias” exists when a data sample that a researcher examines is not representative “of the larger population about which the researcher is trying to draw conclusions.”¹⁹⁷ If selection bias is present, then “conclusions drawn from the sample may not be valid for the

explain such findings. See Hoffman et al., *supra* note 153, at 688; Olson, *supra* note 185, at 790–91. Accordingly, Imber and Gayler’s research provides only marginal support for the claim that reported cases are representative of total filings within a given category. See also Zirkel & Machin, *supra* note 163, at 494 (discussing four other issues with Imber and Gayler’s study).

¹⁹⁴ Samuel R. Gross & Kent D. Syverud, *Getting to No: A Study of Settlement Negotiations and the Selection of Cases for Trial*, 90 MICH. L. REV. 319, 320 (1991) (noting that “[e]veryone seems to agree” that the cases that are tried are not a random subset of all lawsuits).

¹⁹⁵ See Priest & Klein, *supra* note 154, at 4 (“This paper presents a model of the litigation process that clarifies the relationship between the set of disputes settled and the set litigated. . . . From this proposition, the model shows that the disputes selected for litigation (as opposed to settlement) will constitute neither a random nor a representative sample of the set of all disputes.”). Regarding the next level in the pyramid, see Kim et al., *supra* note 176, at 100 n.54 (“Of course, it is also true that filed cases are unrepresentative of all disputes, because potential litigants must first recognize their injuries as grievances and decide to pursue litigation, and their decision is to do so will not be random.”) (citing William L.F. Felstiner et al., *The Emergence and Transformation of Disputes: Naming, Blaming, Claiming* . . . , 15 LAW & SOC’Y REV. 631 (1980)); see also Richard A. Posner, *The Law and Economics of Contract Interpretation*, 83 TEX. L. REV. 1581, 1600 n.48 (2005) (“[T]he contracts that get litigated to the appellate level, which are the contracts scholars mainly discuss, are not a random sample of all contracts.”).

¹⁹⁶ The research comparing published to unpublished cases frequently uses the metaphor of an iceberg: published decisions are the parts above the water’s surface and unpublished orders and lawsuits are the parts below the surface. See, e.g., Mead, *supra* note 171, at 589; Zirkel & Machin, *supra* note 163, at 486–88. The unrepresentative nature of reported decisions can be summarized by mixing the iceberg and pyramid metaphors together: Each level in the pyramid is not representative of the level beneath it, and the top two levels in the pyramid—appellate matters and cases litigated to judgment—are further distorted by the iceberg effect.

For other sources reviewing the literature on whether published cases are representative, see Ringquist & Emmert, *supra* note 188, at 9–10, and Zirkel & Machin, *supra* note 163, at 488–99; see also Hoffman et al., *supra* note 153, at 727–28 (“The underrepresentativeness of opinions is obvious and well known.”).

¹⁹⁷ Taha, *supra* note 175, at 171.

wider population.”¹⁹⁸ Put in technical terms, findings based on a sample infected with selection bias may lack “external validity,” which means the results of the study cannot be “generalized” to a broader population.¹⁹⁹

To avoid the problem of selection bias, scholars normally use random or “probability” sampling.²⁰⁰ For example, in a study of lawsuits filed in a particular jurisdiction, the author could employ a random number generator to select the suits that will be included in the analyzed dataset.²⁰¹ Alternatively, the author might apply systematic sampling, such as picking every fifth case in the population, which is often the functional equivalent of random sampling.²⁰² Unfortunately, as just discussed, reported decisions are not a random subset of other potential categories of study, like litigated judgments, lawsuits, and disputes.²⁰³ This means that empirical studies based on published cases, including those that employ key number coding, probably suffer from some degree of selection bias, limiting the generalizability of any findings.²⁰⁴ Mark Hall and Ronald Wright summarize this problem exceptionally well:

¹⁹⁸ *Id.*

¹⁹⁹ LAWLESS ET AL., *supra* note 19, at 39–40; Korobkin, *supra* note 4, at 1051.

²⁰⁰ See LAWLESS ET AL., *supra* note 19, at 101 (“Random assignment is one of the central features of experimental design that makes causal inferences possible.”); Epstein & King, *supra* note 7, at 110 (“The main reason is that random selection is the only selection mechanism in large-*n* studies that automatically guarantees the absence of selection bias. That is because when we use random sampling we are, by definition, assuring the absence of any association that may exist between selection rules and the variables in our study.”); *id.* at 106 (explaining that a “random probability sample” is “a sample in which each element in the total population has a known (and preferably the same) probability of being selected”); Hall & Wright, *supra* note 3, at 102 (“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.”).

²⁰¹ Hall & Wright, *supra* note 3, at 102.

²⁰² LAWLESS ET AL., *supra* note 19, at 145; Hall & Wright, *supra* note 3, at 102; see also LAWLESS ET AL., *supra* note 19, at 143–48 (discussing various types of random sampling); *id.* at 101–06 (discussing the role of random assignment in experimental design).

²⁰³ See also Edwards & Livermore, *supra* note 161, at 1923 (“Published decisions as a sample of total decisions are far from random[.]”); Berdejo, *supra* note 172, at 274 (“[P]ublished decisions are not a random sample of all decisions rendered by appellate courts.”); Kim et al., *supra* note 176, at 97 & n.44 (“Numerous studies have found that published and unpublished opinions differ in systematic ways.”).

²⁰⁴ Taha, *supra* note 175, at 174 (“Because published opinions are not representative of all cases decided by a court, studies that rely on published opinions to study courts’ decision-making are subject to a selection bias.”); Epstein & King, *supra* note 7, at 106 (“Just as researchers can introduce bias in their studies when they draw unrepresentative samples, the world that creates the sets of *potentially observable data* can also bias inferences if it differs systematically from the target population. . . .

There is slippage at each point in the litigation process: most human interactions do not produce disputes, only some disputes result in legal claims, many claims are settled, and many trial decisions are not appealed. Appellate courts regularly dispose of cases without opinions or decide not to publish some opinions, and computer databases inconsistently include cases that are not officially published. At each juncture, a variety of factors potentially distort what one stage can reveal about the other. These biases can fundamentally threaten the ability to generalize or the validity of a study's findings.²⁰⁵

Note that if the conclusions a researcher intends to draw from a dataset of reported decisions relate only to that type of decision, then it is of no consequence that published rulings are unrepresentative of various broader populations.²⁰⁶ In other words, it is permissible to draw inferences from a sample of reported cases when the broader population under study is *also* constituted by reported caselaw (as long as the sample is randomly generated). A problem exists only when reported decisions are treated as a sample of some *other* type of population, such as all litigated cases or all lawsuits. To illustrate, a scholar might wish to analyze a feature of binding precedent.²⁰⁷ In that situation, it is entirely appropriate to focus exclusively

Among the most prominent examples are studies that base their inferences exclusively on *published* opinions, rather than on the full population of *published* and *unpublished* opinions.”); Hall & Wright, *supra* note 3, at 103 (“[T]he sample frame of all published decisions does not fully reflect what all courts do . . . because most court determinations are unpublished. Whenever the actual cases selected do not fully match the sampling frame that theoretically applies to the questions posed or studied, sampling bias potentially exists. The findings from cases studied may not accurately represent those to which the authors apply their conclusions.”); *see also* Siegelman & Donohue, *supra* note 153, at 1134 (“If appellate cases constitute only a fraction of the population of disputes and potential disputes and if that fraction is not only minuscule, but systematically unrepresentative of the larger population from which it is drawn, generalizing from such cases to larger conclusions about law and/or society is hazardous.”); *but cf.* Frank Cross et al., *Above the Rules: A Response to Epstein and King*, 69 U. CHI. L. REV. 135, 140 (2002) (“Finally, in many studies a focus on published (and the exclusion of unpublished) decisions will not necessarily skew the results in any obvious direction.”).

²⁰⁵ Hall & Wright, *supra* note 3, at 104; *accord* Korobkin, *supra* note 4, at 1054 (“Second, it is well known that cases that result in published opinions are not necessarily a random sample of all cases filed concerning a particular doctrinal topic, much less all the disputes within that doctrinal area. Thus, conclusions appropriately drawn about published decisions may not be generalizable to all decisions, much less all litigation in a particular substantive area, or all disputes.”).

²⁰⁶ Hall & Wright, *supra* note 3, at 104 (“No concern arises if the researcher defines the research questions in terms that match the population of cases actually sampled.”).

²⁰⁷ *See, e.g.*, Keckler, *supra* note 82, at 108–09, 131–32, 146 n.66 (empirical study of how Illinois courts apply *stare decisis*); *see also* White, *supra* note 56, at 1275 (using a dataset of reported decisions to conclude that the Uniform Commercial Code “has not caused a radical change in the type of sales litigation that appears in the reported decisions”).

on reported, appellate rulings.²⁰⁸ However, most empirical research about the legal system seeks to draw inferences about categories that go beyond published opinions. When that is true of a study employing the Key Number System as the coding tool, the author must be cautious in deriving conclusions about populations that encompass more than reported caselaw, for the reasons just discussed.

While the selection bias inherent in datasets of published decisions is an important weakness of studies using key numbers, it does not justify abandoning this type of work. It is well recognized that research lacking features necessary for scientific validity, such as random assignment, can still provide valuable empirical insights.²⁰⁹ “[S]ocial science data does not need to be perfect; reasonable approximations are good enough for both government work and empirical work.”²¹⁰ That is especially so when the alternative to a study employing key numbers is not a better project but no project at all. And in many circumstances, compiling a dataset that is superior to reported decisions is cost-prohibitive or truly impossible.²¹¹ Even when a competing project that reduces or avoids selection bias is merely more difficult than a key number study, the resources saved in

²⁰⁸ Hall & Wright, *supra* note 3, at 104–05 (“For instance, unpublished opinions are irrelevant to the study of precedential law rather than the generalized behavior or attitudes of judges, so excluding them requires no justification.”); Kim et al., *supra* note 176, at 96 (“Focusing only on published decisions may make sense when asking, for example, how formal legal doctrine has evolved.”).

²⁰⁹ See, e.g., LAWLESS ET AL., *supra* note 19, at 115–17 (explaining that quasi-experiments, which approximate traditional experiments but lack certain important features, such as random assignment, still can provide useful empirical findings despite the fact that quasi-experiments “leave open the possibility of an alternative explanation of any observed effect”).

²¹⁰ Hall & Wright, *supra* note 3, at 97.

²¹¹ See, e.g., *id.* at 94, 97 (explaining that in studies of the factors that determine the outcomes of cases, the imperfect data contained in judicial opinions often “must suffice because observing actual behaviors and gauging true attitudes would be impossible or cost-prohibitive”); Hooker, *supra* note 114, at 9 (using key number searches to measure trends in the quantity of education-related litigation because “[t]here is no reasonable alternative source of data”). Hooker elaborated as follows:

Theoretically, a complete and accurate account of all litigation involving teachers could be obtained from the records and archives of the 15,000 school districts in the nation. This method of gathering data for this study was considered and discarded after conducting a pilot study in Minnesota. Incomplete records in school districts, school district consolidations, reporting errors by school officials, incomplete responses to a questionnaire and numerous additional logistical problems made West’s American Digest System an easy favorite data base for this study.

conducting the latter will often justify following the less scientifically correct approach.²¹²

To elaborate, studies using the Key Number System have substantially the same strengths and weaknesses in the area of generalizability as other research grounded on reported cases. Critically, empirical assessments of published opinions are increasing in number and social scientists have long believed that there is significant value in scholarship based on these materials.²¹³ Moreover, it is common for scholars to draw inferences about broader populations from samples of reported decisions, including when the coding is performed via key numbers. For example, Farber and Matheson used a dataset of 222 published cases involving promissory estoppel to reach conclusions about promissory estoppel litigation in general.²¹⁴ Hillman did the same with 363 reported opinions on that subject.²¹⁵ And Bratton and Wachter used cases “keyed by West as involving preferred

²¹² See Lee Epstein & Gary King, *A Reply*, 69 U. CHI. L. REV. 191, 207 (2002) (observing that “perfectionism in [empirical] method at the expenses of other goals is both inappropriate and unnecessary”). Hall and Wright offer the following example. Studies of the factors that determine the outcomes of cases frequently rely upon the facts as presented in court decisions. Hall & Wright, *supra* note 3, at 94. But the statements of facts in judicial opinions are not always accurate. *Id.* at 95. Hall and Wright nonetheless conclude that the labor saved in utilizing such opinions is one reason that it is appropriate to rely upon a resource that is only an approximation of reality. *Id.* at 97.

²¹³ See Hall & Wright, *supra* note 3, at 98–99 (observing that social scientists have long thought judicial opinions are extremely valuable source of information on the legal system and that “[m]odern day legal researchers tend to agree, judging from how many have devoted considerable effort to using this data source”); see also Taha, *supra* note 175, at 173 (“Much empirical research of civil litigation has used published opinions as the source of data.”); see generally Hall & Wright *passim*.

²¹⁴ Farber & Matheson, *supra* note 41, at 904, 907. For additional details regarding this study, see *supra* notes 41–43 and accompanying text.

²¹⁵ Hillman, *supra* note 44, at 580–82, 619; see, e.g., *id.* at 581 (“[T]his article is the first comprehensive empirical study that demonstrates promissory estoppel’s limited role.”); *id.* at 619 (“[B]ut the purpose of this Article is to *show* that courts do not enforce promises without bargain or reliance.”) (emphasis in original). For additional details regarding this study, see *supra* notes 44–47 and accompanying text. For some additional studies from the contracts literature where scholars used a dataset of published cases to generalize about broader populations, see Crystal, *supra* note 56, at 299 (using a dataset of 262 reported decisions to answer three questions regarding the litigation of relational contracts under Article Two of the Uniform Commercial Code; each of the questions was phrased in terms that go beyond published opinions); DiMatteo & Rich, *supra* note 56, at 1069, 1092–93 (using a dataset of 187 primarily reported, federal cases to answer four questions regarding the doctrine of unconscionability; once again, each of the questions was phrased in terms that go beyond published opinions); McChesney, *supra* note 56, at 171–73, 185 (using a dataset of 134 reported decisions to conclude that in tortious interference with contract cases courts are focused on protecting the property interests of the nonbreaching party rather than with efficient and inefficient breaches).

stock” to support the broad proposition that “Delaware courts have emerged as the dominant arbiters of preferred stock disputes.”²¹⁶ None of this is meant to suggest that the frequency of studies generalizing from reported caselaw negates the problem of selection bias. Rather, it is intended to point out that employing the Key Number System is consistent with protocols long-used and valued by other researchers.

Note further that legal scholarship based on alternative sources of information also frequently suffers from selection bias. For example, in Macaulay’s study of the contracting practices of business people, most of the subjects he interviewed worked at businesses or law firms in a single state (Wisconsin), and numerous industries were not represented among the interviewees.²¹⁷ As a result, Macaulay recognized that “the likelihood of error because of sampling bias may be considerable.”²¹⁸ This problem is widespread. To illustrate, in 2002, Russell Korobkin wrote an article analyzing the empirical contracts scholarship produced over the prior fifteen-years.²¹⁹ He concluded that *none* of the twenty-seven studies he reviewed used “data that makes up a truly representative sample of the population (i.e., contracting parties or contract disputes generally) about

²¹⁶ See Bratton & Wachter, *supra* note 107, at 1821 & n.14. This study is discussed further at *supra* notes 107–08 and accompanying text. For other examples of empirical research employing key numbers where the authors generalized their findings to broader populations, see Lisby, *supra* note 113, at 466–67 & nn.252–54 (using cases included in the American Digest under the key numbers relevant to criminal libel to show a decline in “[p]rosecutions for the crime of libel”); Wines, *supra* note 113, at 86–87 (using digested cases as a proxy for total litigation regarding the tort doctrine of “frolic and detour”); Zirkel & Richardson, *supra* note 109, at 789 (using cases keyed with education key numbers to draw conclusions regarding total levels of education litigation and levels of litigation by education subcategory); Snyder, *supra* note 75, at 547, 576–77 (contending that how often each digest topic appears in all Montana Supreme Court and Ninth Circuit Court of Appeals cases decided from 1945 through 1997 is relevant to the question of which areas of law are most litigated).

²¹⁷ Macaulay, *supra* note 35, at 55–56.

²¹⁸ *Id.* at 56; see also Korobkin, *supra* note 4, at 1052–53 (“Macaulay made no attempt to ensure that this sample was representative of the world of commercial contracting parties in terms of industry, size of business, etc.”).

²¹⁹ Korobkin, *supra* note 4, at 1033.

which the author would like to suggest implications.”²²⁰ Surveys of empirical work in other fields of law have reached similar conclusions.²²¹

Korobkin actually went a step further and argued that the problem of selection bias “cannot be solved merely by researchers paying closer attention to methodological considerations, for two reasons.”²²² First, resource limitations make it exceptionally difficult for a scholar to gather a dataset that is properly representative across all relevant dimensions.²²³ For example, a number of authors contend that empirical studies of federal trial courts should focus on dockets and court files rather than published decisions.²²⁴ But docket studies usually suffer from some type of geographical selection bias.²²⁵ And obtaining a representative sample from across the nation will probably be cost prohibitive in most situations. Second, and more critically, Korobkin contends that “it is usually impossible to design a data set that is perfectly representative of the larger target population in every relevant way.”²²⁶ And even if creating such a

²²⁰ *Id.* at 1051, 1064–66; *see also id.* at 1053 (explaining that all of the studies reviewed in the article that were based on surveys involved unrepresentative samples because “not everyone asked to participate agreed, raising the concern that the responses of people who agree to take the time to participate in a research study might be systematically different from those who would not take the time”); *see generally id.* at 1051–56 (discussing the generalizability of empirical findings).

²²¹ *See, e.g.,* David Freeman Engstrom, *The Twiqbal Puzzle and Empirical Study of Civil Procedure*, 65 STAN. L. REV. 1203, 1214–17 & n.45 (2013) (discussing sampling bias in empirical research regarding the impact of recent changes to federal pleading rules); Nadine Frederique et al., *What is the State of Empirical Research on Indigent Defense Nationwide? A Brief Overview and Suggestions for Future Research*, 78 ALB. L. REV. 1317, 1325–31 (2014–15) (concluding that empirical research on criminal defense attorney effectiveness has produced “few generalizable results”); Bryan G. Garth, *Observations on an Uncomfortable Relationship: Civil Procedure and Evidence Research*, 49 ALA. L. REV. 103, 108–09 (1997) (explaining that “randomized experiments in civil justice research are quite rare” and thus many studies suffer from selection bias).

²²² *Id.* at 1054.

²²³ *Id.*

²²⁴ *See* Hoffman et al., *supra* note 153; Kim et al., *supra* note 176; Schlanger & Liebman, *supra* note 153.

²²⁵ *See, e.g.,* Hoffman et al., *supra* note 153, at 708, 734–36 (empirical study of federal trial court judges based upon the District of Maryland, the Northern District of California, the Southern District of New York, and the Eastern District of Pennsylvania; the authors acknowledged that the latter two districts “are not ordinary” and that this “confound[s] our results”).

²²⁶ Korobkin, *supra* note 4, at 1054. Korobkin continues: “In studies of actual contracting behavior of commercial parties, for example, it would be impossible to identify all the features of contracting parties that might affect their behavior.” *Id.* These factors include business size, industry type, geographical region, whether the business is new or established, the number of locations a business has, who in the business makes the contracting decisions, and various personal characteristics of the

sample were feasible, “the results would be open to the conjecture that, while applicable on an aggregate level, they might not be applicable” to numerous particular subgroups of the overall population.²²⁷ Whether Korobkin is correct or not that selection bias is unavoidable, the following is undeniable: When it comes to selection bias, key number studies have many companions in guilt.

Ultimately, the problem of generalizability is not a reason to forego empirical research using key numbers, cases available on Westlaw and Lexis, the lawsuits filed in a particular jurisdiction, or other non-representative samples.²²⁸ As noted previously, all empirical scholarship is imperfect.²²⁹ Every research protocol has advantages and disadvantages, requiring scholars to make difficult trade-offs between external validity, available resources, and other methodological and practical factors.²³⁰ And studies based on non-random datasets, which are quite common,²³¹ often possess significant value because they contribute to the collective enterprise of knowledge acquisition regarding the legal system or other social phenomena.²³² Instead, concerns about selection bias are a reason to

principal decision maker. *Id.* at 1054–55. “No study could conceivably be representative along all of these dimensions, much less the untold number of other potentially relevant factors.” *Id.* at 1055; *see also, e.g.,* Geis, *supra* note 4, at 462 n.57 (explaining that the EDGAR database of material contracts filed with the SEC by public firms is subject to two types of selection bias).

²²⁷ *Id.* at 1055.

²²⁸ Korobkin, *supra* note 4, at 1055 (“The generalizability problem, then, should not discourage would-be empirical researchers”); *see also* Gregory C. Sisk & Michael Heise, *Judges and Ideology: Public and Academic Debates About Statistical Measures*, 99 NW. U. L. REV. 743, 792 (2005) (making the same point about methodological problems in general).

²²⁹ *See supra* text accompanying note 210.

²³⁰ LAWLESS ET AL., *supra* note 19, at 46; Epstein & King, *supra* note 212, at 207–08; Jack Goldsmith & Adrian Vermeule, *Empirical Methodology and Legal Scholarship*, 69 U. CHI. L. REV. 153, 154 (2002).

²³¹ LAWLESS ET AL., *supra* note 19, at 148 (noting that when random (i.e., “probability”) sampling is impossible or difficult, “it is common for a variety of” nonrandom (i.e., “nonprobability”) sampling techniques to be used, including “convenience sampling—sampling units that are readily available rather than those that are randomly selected”); *see also id.* at 148–49 (discussing various types of nonrandom sampling); Kevin H. Smith, *External Validity: Representativeness and Projectability in the Probative Value of Sample Surveys*, 39 WAYNE L. REV. 1433, 1503–04 (1993) (identifying the circumstances in which non-random sampling is justified, including (1) when insufficient resources are available to obtain a random sample, and (2) when trying to determine whether a problem exists that might justify a considerably more expensive study involving probability sampling).

²³² Korobkin, *supra* note 4, at 1055 (explaining, *inter alia*, that when a study “is able to reach strong conclusions about its limited data set . . . such a study undoubtedly makes a valuable contribution to scholarship”); A. Lynn Phillips et al., *What’s Good in Theory May Be Flawed in Practice: Potential*

carefully acknowledge any such weakness and exercise caution in extrapolating from one's sample to broader populations.²³³ As Hall and Wright explain in the context of empirical work based on judicial decisions,

[a]ll empirical studies are imperfect, especially observational (nonexperimental) social science studies. The goal in selecting cases is not a perfect match between sample frame and research conclusions, but only a reasonable connection between the two. Inevitable imperfections in case selection methods often will not seriously threaten the entire validity of the study's findings. It usually suffices to acknowledge limitations fairly briefly.²³⁴

The unrepresentative nature of reported decisions is not the only limitation involved in using the Key Number System for data coding. There are multiple others. I discuss ten such limitations over the next few pages. Nine of these concern the precise manner in which West writes and classifies headnotes for cases. The tenth concerns the search capacity of Westlaw. Note that I also address a number of issues that are specific to my study protocol later in this article.²³⁵

First, West does not always digest cases accurately. Classifiers sometimes code a headnote with the wrong key number.²³⁶ Second, and

Legal Consequences of Poor Implementation of a Theoretical Sample, 9 HASTINGS BUS. L.J. 77, 89 (2012) (“Sometimes there are good reasons for a doing a nonprobability sample—if a researcher does not know the characteristics of the target population or it would be prohibitively expensive to do a probability sample, he might still learn valuable information from a nonprobability sample.”); see also Smith, *supra* note 231, at 1502 (acknowledging that, in some contexts, as the number of studies on a topic using non-probability sampling and reaching the same conclusion grows, the likelihood that the results are representative of the relevant larger population increases).

²³³ See Epstein & King, *supra* note 212, at 208 (“The sign of good science is *not* that every problem be fixed in every project, since that is impossible and trying to do so would mean we would not accomplish much of anything, but rather that the problems be noted and the uncertainty in substantive conclusions be responsibly and honestly reported.”). When using a non-random sample, scholars should also consider explaining why they chose to employ a less scientifically sound methodology.

²³⁴ Hall & Wright, *supra* note 3, at 105; see also Hoffman et al., *supra* note 153, at 686–87 (noting that most authors who conduct empirical research based on published opinions “are careful to recognize the opinions might be unrepresentative of how trial courts resolve legal problems”) (collecting authorities).

²³⁵ See *infra* notes 476–80 and accompanying text; Part V.C.1.

²³⁶ Michael H. Fox & Robert A. Mead, *The Relationship of Disability to Employment Protection Under Title I of the ADA in the United States Circuit Courts of Appeal*, 13 KAN. J.L. & PUB. POL'Y 485, 496 n.106 (2004) (noting that there is a “slight potential for a relevant case to not be categorized under the appropriate topic and key number”); Spencer L. Simons, *Navigating Through the Fog: Teaching Legal Research and Writing Students to Master Indeterminacy through Structure and Process*, 56 J. LEG. EDUC. 356, 362 (2006) (noting that “cases are occasionally misclassified by digest editors”); see Daniel P. Dabney, *The Curse of Thamus: An Analysis of Full-Text Legal Document Retrieval*, 78 LAW

related, there may be some variation in the level of detail that classifiers use in coding opinions. One attorney-editor might tag a decision with many key numbers that represent highly specific points of law, while another would code the same case with only a few key numbers that denote more general legal concepts.²³⁷ However, “the reliability and accuracy of the West editorial staff” has long been considered one of the principal strengths of the American Digest System.²³⁸ Errors and disparate classification practices should thus be rather uncommon.²³⁹ And, as previously explained, West attorneys (1) lack certain biases that can infect coding performed by authors, and (2) are probably superior coders to law students.²⁴⁰

Third, key numbers are sometimes included in cases even though the associated topic was not litigated. This typically occurs when courts explain general principles that are related to the contested issues.²⁴¹ Such “over-coding” can impact the reliability of a dataset created via the American

LIB. J. 5, 14 (1986) (“This short review of ideas in indexing shows that the indexing process is prone to many sorts of errors and uncertainties. . . . Finally, like any human enterprise, it is not always done as well as it might be.”); Geis, *supra* note 4, at 479–80 (offering an example of how coding mistakes by West attorney-editors might have biased the results of an empirical study that used the Key Number System to collect opinions); *see also* LAWLESS ET AL., *supra* note 19, at 130 (noting that one issue with secondary data analysis “is the reliability of the data that is recorded”).

²³⁷ Shapiro, *supra* note 3, at 528 n.213 (“On the other hand, the level of detail provided appears to vary greatly from case to case, with some cases identified by many, often repetitive keys, and others identified by only the most general.”). There may also be differences in headnote writing practices. I have at times found that points of law in a case were not abstracted. I suspect that variations in headnote drafting are most common with respect to (1) discussions of related legal principles not directly relevant to the matter before the court, and (2) long summaries of basic principles set forth prior to the specific governing rules and authorities.

²³⁸ Berring, *supra* note 87, at 33.

²³⁹ *See id.* at 34 (“It is difficult to assess the extent of this problem [misclassification]. My impression is that it was not severe, West’s reputation for accuracy was well deserved.”). I contacted West about this issue. A representative explained that West uses various quality control procedures. But unfortunately the company does not have specific reports or studies that they could share with me. E-mail from Scott Augustin, Senior Director, Corporate Affairs, Thomson Reuters, to Author (Oct. 30, 2015, 9:45 CST) (on file with author).

²⁴⁰ *See supra* notes 138–42 and accompanying text.

²⁴¹ *See, e.g.*, *City of Fort Wayne v. Consol. Elec. Distr., Inc.*, 998 N.E.2d 733, 734 (Ind. Ct. App. 2013) (opinion contains a headnote with key number 95k158, a contract interpretation classification, because the court stated a general principle about contract interpretation in the course of addressing a statutory interpretation question; there was no contract at issue in the case).

Digest System.²⁴² For example, suppose a scholar is using decisions tagged with various key numbers to measure how often a particular type of case is brought.²⁴³ Any over-coding involving the selected key numbers will cause Westlaw search results to exaggerate the incidence of the measured type of lawsuit.

What is the prevalence of over-coding? Unfortunately, that is unknown. But I suspect that for most empirical projects involving large samples, either (1) the over-coding will balance out, or (2) the frequency of over-coding will be too small to have a significant impact on the ultimate findings. Let me briefly elaborate on what I mean by “balance out.” In my experience, the level of key number over-coding generally does not differ by subject or jurisdiction. Thus, in an empirical study comparing two sets of cases, the amount of over-coding normally should be the same in each set—i.e., it should “balance out.” If this is correct, then the over-coding will not bias the study results.

However, there are reasons to believe that over-coding might vary by topic or state in some circumstances. Over-coding could thus be a concern for certain empirical projects. Accordingly, it is advisable that researchers assess the potential for over-coding problems early in their research.²⁴⁴ Here is an example that illustrates the issue. New York reported decisions are frequently very short. Over-coding is less likely in short opinions because such cases typically omit the extensive discussions of background legal principles that are the primary cause of over-coding. As a result, a key number study comparing New York to another jurisdiction where longer opinions are more common might be infected with over-coding bias.²⁴⁵

Fourth, there is considerable overlap between many of the digest topics and subtopics.²⁴⁶ That creates room for judgment when West’s staff is

²⁴² See Hall & Wright, *supra* note 3, at 106 (“Supplemental techniques are needed to screen out cases that mention a topic of interest only in passing or those that decide an issue on technical or procedural grounds irrelevant to the study.”).

²⁴³ See, e.g., Bratton & Wachter, *supra* note 107, at 1821 & n.14. This study is discussed *supra* at notes 107–08 and accompanying text.

²⁴⁴ See *infra* notes 472–76 and accompanying text (setting forth how I assessed the possibility of over-coding bias in my contract interpretation study).

²⁴⁵ My thanks to Royce Barondes for pointing out this possibility.

²⁴⁶ See Snyder, *supra* note 75, at 548 (noting that “some digest topics . . . closely shade into one another”); Zirkel & Richardson, *supra* note 109, at 769 (“For example, the West’s [sic] key number system . . . is subject to inevitable imprecision with regard to the identification, placement, and addition

making classification decisions. As a result, for numerous key number topics, there will be cases with headnotes that could have been classified under that topic, but which were instead classified under another, related topic.²⁴⁷ This might be thought of as a type of “under-coding.” To illustrate, scholars who have used key number searches to measure the quantity of education litigation frequently note that cases concerning that subject are sometimes not coded with any key numbers that directly implicate education. Instead, they are classified with key numbers from other topics such as Constitutional Law and Civil Rights.²⁴⁸ One study attempted to measure the extent of this problem and found that, of all the cases listed in the Education Reporter over a two-year period, only seven lacked any education-related key numbers.²⁴⁹ This suggests that the number of under-codings is often too small to warrant concern.²⁵⁰ But that is not always true. For example, there are more than 10,000 decisions involving the construction of insurance contracts that are coded with contract interpretation key numbers from the topic Insurance (217) but contain no interpretation key numbers from the topic Contracts (95).²⁵¹ This means that an empirical study of interpretation caselaw that uses only key numbers categorized under Contracts will miss numerous insurance-related decisions on the subject.²⁵²

Fifth, the American Digest System “does a poor job of classifying cases in certain areas.”²⁵³ For example, Fritz Snyder argued that because

of categories. And as with any other indexing system, some cases fit within more than one subject category.”).

²⁴⁷ See Snyder, *supra* note 75, at 548 (explaining that researchers sometimes have to check multiple digest topics in order to find all cases regarding a particular subject).

²⁴⁸ See Irene Gavin & Perry A. Zirkel, *An Outcome Analysis of School Employee-Initiated Litigation: A Comparison of 1977–1981 and 1997–2001 Decisions*, 232 WEST’S EDUC. L. REP. 19, 22–23 (2008) (observing that some K-12 litigation is coded with key numbers from topics other than Schools (345), such as Civil Rights (78), Constitutional Law (92) and Labor Relations (232A)); Zirkel & Johnson, *supra* note 114, at 8 & n.26 (“[P]revious studies . . . revealed that some of the K-12 decisions only have key numbered head notes for categories outside 345 (Schools).”).

²⁴⁹ Hooker, *supra* note 114, at 9.

²⁵⁰ *Id.*

²⁵¹ To see this, run the following search on Westlaw in the database containing all state and federal caselaw: 217XIII(G) % 95II(A). The result contains over 10,000 cases.

²⁵² For a discussion of this issue in the context of my contract interpretation study, see *infra* notes 468–71 and accompanying text.

²⁵³ Snyder, *supra* note 75, at 549.

Evidence (157) and the various topics related to civil procedure²⁵⁴ are not organized by rule, it is easier to conduct research about these subjects using other sources.²⁵⁵ Likewise, much of my practice before entering academia concerned antitrust law, and I found Monopolies (265), which previously covered this field, to be rather unhelpful in my work.²⁵⁶ Topics that are not useful in traditional legal research are likely to be even less suitable for empirical studies.

Sixth, some of the digest topics and subtopics are exceptionally broad.²⁵⁷ Indeed, many individual key numbers cover multiple, distinct legal subjects. For example, several of the key numbers relating to contract interpretation also apply to the construction of other types of legal documents, such as wills.²⁵⁸ Fortunately, I suspect that this type of “overbreadth”—as with over-coding—will balance out in studies employing large samples. But classifications that suffer from overbreadth are probably not appropriate for empirical work in at least some circumstances.²⁵⁹

Seventh, because changes to the American Digest System come at a slow pace, certain topics do not keep up with developments in their associated legal field.²⁶⁰ This potentially limits the value of key numbers for empirical research on newer issues.

The third through seventh limitations entail that when using key numbers for data coding, scholars must develop some expertise with the

²⁵⁴ For example, Federal Civil Procedure (170A), Federal Courts (170B), Pleading (302), and Pretrial Procedure (307A). WEST’S ANALYSIS OF AMERICAN LAW, *supra* note 69, at 751, 778, 1356, 1376.

²⁵⁵ Snyder, *supra* note 75, at 549–50.

²⁵⁶ Note that since I left practice, West moved antitrust law out of Monopolies (265) and into a separate topic called Antitrust and Trade Regulation (29T). See WEST’S ANALYSIS OF AMERICAN LAW, *supra* note 69, at 54, 1235.

²⁵⁷ See Snyder, *supra* note 75, at 577 (“Some of the 404 digest topics are confusing. Some are extremely broad and, in fact include other, narrower topics. Some are very narrow.”).

²⁵⁸ See WEST’S ANALYSIS OF AMERICAN LAW, *supra* note 69, at 706–10 (setting forth the key numbers under topic 157 (“Evidence”) which govern “Parol or Extrinsic Evidence Affecting Writings”). To illustrate, 157k452 concerns latent ambiguities in multiple types of legal documents, including contracts.

²⁵⁹ See *infra* notes 472–76 and accompanying text (setting forth how I addressed the overbreadth issue in my contract interpretation study).

²⁶⁰ Simons, *supra* note 236, at 362; see also Lee F. Peoples, *The Death of the Digest and the Pitfalls of Electronic Research: What is the Modern Legal Researcher to Do?*, 97 LAW LIBR. J. 661, 666 (2005) (observing that digest topics are changed at a “conservative pace”).

pertinent digest topics. This will enable them to (1) determine whether their proposed study is feasible, (2) construct appropriate Westlaw searches, and (3) provide any necessary qualifications when presenting their findings.

Eighth, West frequently makes changes to the key number topics. This includes both restructuring individual topics and moving material from one topic to another.²⁶¹ Such alterations can make it difficult or even impossible for an author to run searches in Westlaw that are identical or comparable to those run by prior authors. This limits the replicability of earlier work and will sometimes prevent a researcher from crafting studies that make small adjustments to previously-employed protocols in an attempt to test an hypothesis using a slightly different research design.²⁶²

Ninth, the American Digest System does not code for numerous elements of judicial opinions. When a headnote is classified with a given key number, all this conveys is that a particular legal subject was discussed

²⁶¹ For more, see the sources cited in notes 92–94 *supra*.

²⁶² See Zirkel & Johnson, *supra* note 114, at 7 (“West’s reclassification of cases from one key number subcategory to another also contributes to changes in the numbers from one data collection to another, especially with approximately ten years intervening between the tabulations. Similarly, the shift in key numbers for certain subcategories also contributes to the variance.”).

To elaborate, when West reorganizes a segment of the Key Number System, the changes are *retroactively* applied to *all* cases contained on Westlaw. E-mail from Scott Augustin, *supra* note 239. But to aid researchers, West generally includes references identifying the original key number classifications for amended headnotes. *Id.* These references are set forth directly in the revised headnotes and begin with the word “formerly.” *See id.* For example, Insurance (topic 217) was restructured in 1998; a completely new set of key numbers was adopted for that topic. The key number classifications for all Insurance headnotes digested prior to the change were retroactively updated online, while the now-defunct key numbers assigned to each case were preserved via the “formerly” references. (They were also preserved in the print copies of the reporters.) Note that cases published *after* a switch are classified only with the revised key numbers. And thus more recent insurance cases contain only the newer classifications. (To skim through opinions reflecting the amendments to the Insurance topic, run the following search on Westlaw for all state and federal cases: (TO(217) /p (formerly /1 217k!)) & DA(aft 05/01/1998) & DA(bef 08/01/1998). The switch took place in June of 1998.)

Suppose a scholar conducted an empirical study using the Insurance key numbers in early 1998 that concerned the time period 1996 through 1997. A second researcher today could run identical or modified searches using the old classifications for any time period prior to June, 1998, thanks to the “formerly” references. But if the second researcher wants to replicate the original study in a more recent time period, that will often be impossible—or at least extremely difficult—for two reasons. First, the insurance cases decided since the topic restructuring in 1998 only have the new key numbers. Second, in many if not most circumstances, the topic was changed too dramatically to craft searches using the new key numbers that are truly comparable to the searches run in the original study with the older key numbers.

by the court—and, for multi-judge courts, that it was addressed by the majority.²⁶³ Critically, many empirical studies of caselaw concern aspects that go beyond the subject matter of the litigation. The author wishes to analyze “the substance of judicial reasoning, as expressed through the legal and factual content of written opinions.”²⁶⁴ Such information typically can be ascertained only via a careful reading of the decision.²⁶⁵ In that circumstance, key numbers will obviously not be helpful.²⁶⁶ But other studies focus on more basic information, such as subject matter, parties, and outcomes.²⁶⁷ For this type of work, key numbers can prove quite valuable.

Tenth, Westlaw caps the number of characters (including spaces) that may be included in a search of its databases at 600.²⁶⁸ That restriction will sometimes prevent a scholar from using the digest system to conduct studies that necessitate including a large number of key numbers in single search.²⁶⁹

The ten limitations just discussed, even when taken together, are no more reason to avoid empirical research based on the American Digest System than the problem of generalizability. Accordingly, given the

²⁶³ West does not write headnotes for concurrences and dissents unless “a point of law within the concurrence or dissent commands a majority on that point of law. This is not a frequent occurrence, and when it does happen it is generally in plurality decisions.” E-mail from Steve Hanson, Reference Attorney, Thomson Reuters, to Author (June 19, 2015, 16:51 CST) (on file with author) (apparently quoting an e-mail from Patty Larson, Director of Judicial Editorial, Thomson Reuters, to Steve Hanson). Thus, concurrences and dissents are almost never tagged with key numbers.

²⁶⁴ Hall & Wright, *supra* note 3, at 72–73.

²⁶⁵ *Id.* at 71 n.29 (in this study of case coding research, the authors limited their analysis to “projects that coded for legal, factual, analytic, or linguistic elements of legal decisions that could be gleaned only by a close reading of the opinions, rather than, for instance, information available in a digest or abstract of the decision”).

²⁶⁶ See Cross, *supra* note 103, at 200–01 (explaining that the author’s study using the Key Number System to measure an increase in references to pragmatism by circuit courts only measured the “tip of the iceberg of pragmatism’s usage” in part because “many applications of pragmatism are outside of Westlaw’s coding measure”). For additional details regarding this study, see *supra* notes 103–06 and accompanying text.

²⁶⁷ See Hall & Wright, *supra* note 3, at 72 (in this study of case coding research, the authors “excluded projects using a method we call ‘docket analysis,’ which codes only for information about cases—such as subject matter, parties, and basic outcomes—that could be obtained from docket sheets or brief abstracts”) (identifying two examples); *id.* at 82 n.72 (“Some [studies] code only for case subject matter and outcome. . . . Others simply code for which authorities the opinion cites.”) (identifying four examples of the first type of study and five examples of the second type).

²⁶⁸ E-mail from Daniel Reynolds, Reference Attorney, Thomson Reuters, to Author (Aug. 5, 2015, 17:54 CST) (on file with author).

²⁶⁹ For more on this issue, see *infra* note 449.

advantages identified above, using key numbers in empirical work will often be worthwhile.

Of course, whether a key number protocol is suitable for a particular project will depend on a variety of theoretical and practical factors. For example, on the one hand, key number studies can employ much larger data sets than many other types of empirical work because the researcher need not spend time reading the cases that make up the data set. But key number classifications only capture a narrow range of information about judicial opinions and may lack the necessary level of precision due to over-coding and under-coding. On other hand, carefully reading cases or comparable materials enables a scholar to gather richer data and can address many precision concerns. But completing such reading requires the expenditure of considerably more resources and generally entails using a much smaller sample size. In deciding between a key number study and an alternative methodology, scholars must consider these types of trade-offs.

To further demonstrate both how to deploy the Key Number System as a data collection and coding device and the value of a project that does so, Part V of this article sets forth an empirical study of contract interpretation using key numbers. But to place that study in context, Part IV first presents an overview of the law and policy of contract interpretation.

IV. CONTRACT INTERPRETATION

A. The Textualist and Contextualist Approaches

Interpretation is the process of determining the meaning of contractual language.²⁷⁰ The goal of contract interpretation is to ascertain the intentions of the parties at the time the agreement was formed.²⁷¹ But accomplishing

²⁷⁰ RESTATEMENT (SECOND) OF CONTRACTS § 200 (1981); ALLAN E. FARNSWORTH, FARNSWORTH ON CONTRACTS § 7.7, at 439 (4th ed. 2004).

²⁷¹ BURTON, *supra* note 9, § 1.1, at 1 (“American courts universally say that the primary goal of contract interpretation is to ascertain the parties’ intention at the time they made their contract.”); *accord* RICHARD LORD, 11 WILLISTON ON CONTRACTS § 30:2, at 17–18 (4th ed. 2012); JOSEPH M. PERILLO, CALAMARI AND PERILLO ON CONTRACTS § 3.13, at 136 (6th ed. 2009); *but see* Val D. Ricks, *The Possibility of Plain Meaning: Wittgenstein and the Contract Precedents*, 56 CLEV. ST. L. REV. 767, 807 (2008) (distinguishing between the intention of the parties and the meaning of words).

this task can be difficult because party intent is often unclear and disputed²⁷² and contracts frequently contain ambiguous language.²⁷³

There are two general approaches to contract interpretation set forth in the caselaw. These approaches have multiple names, but I shall use the labels “textualist” and “contextualist.”²⁷⁴ Under textualism, interpretation focuses principally on the language of the parties’ agreement.²⁷⁵ The locus of contextualist interpretation is broader. While adherents of contextualism grant critical weight to the words set forth in the parties’ compact,²⁷⁶ contextualist interpretation emphasizes reading contractual language *in context*.²⁷⁷ Thus, contextualist authorities focus on both the contract’s express terms and extrinsic evidence²⁷⁸—i.e., evidence outside the four

²⁷² See George M. Cohen, *Interpretation and implied terms in contract law*, 6 ENCYCLOPEDIA OF LAW AND ECONOMICS 125, 130 (Gerrit De Geest ed., 2011) (discussing the uncertainty of party intent).

²⁷³ Contractual ambiguities exist for numerous reasons. See FARNSWORTH, *supra* note 270, § 7.8, at 443–44 (setting forth a list). For example, parties typically lack the knowledge and foresight necessary to anticipate every contingency that might be worth addressing in their agreement. BURTON, *supra* note 9, § 1.2.2, at 12–13. Likewise, the stakes in most transactions do not justify the costly and protracted negotiations that would be needed to carefully address all of the issues known to the parties. *Id.* at 13. Finally, and perhaps most fundamentally, language is simply an imperfect medium for expressing ideas. CHARLES KNAPP ET AL., *PROBLEMS IN CONTRACT LAW* 374 (7th ed. 2012).

²⁷⁴ See, e.g., Schwartz & Scott, *supra* note 9, at 928 (using the terms “textualist” and “contextualist” to describe the two schools of interpretation); Cohen, *supra* note 272, at 131, 137 (same). For other approaches to labeling the two schools, see FARNSWORTH, *supra* note 270, § 7.12, at 465 (“restrictive” interpretation versus “liberal” interpretation); James W. Bowers, *Murphy’s Law and the Elementary Theory of Contract Interpretation: A Response to Schwartz and Scott*, 57 RUTGERS L. REV. 587, 589–90 (2005) (“formalist” interpretation versus “contextualist” interpretation); see also Grumman Allied Industries, Inc. v. Rohr Industries, Inc., 748 F.2d 729, 733–34 (2d Cir. 1984) (“classical” interpretation versus “modern” interpretation).

²⁷⁵ See *Grumman Allied Industries*, 748 F.2d at 733–34 (“Adherents of the classical approach, animated by a belief that a contractual agreement manifests the intent of the parties in a completely integrated form, favor the construction of contracts by reference to explicit textual language.”).

²⁷⁶ Bowers, *supra* note 274, at 592 (“Words the parties expressly use play decisive roles in interpretation questions [for contextualist courts].”); see, e.g., RESTATEMENT (SECOND) OF CONTRACTS § 214, cmt. b (1981) (“[T]he words of an integrated agreement remain the most important evidence of intent.”).

²⁷⁷ See *Grumman Allied Industries*, 748 F.2d at 734 (“Modern . . . interpretation . . . seems to derive from the premise that a contextual inquiry is a necessary and proper prerequisite to an understanding of the parties’ intent.”).

²⁷⁸ See *Casey v. Semco Energy, Inc.*, 92 P.3d 379, 383 (Alaska 2004) (“[E]xtrinsic evidence is always admissible on the question of the meaning of the words of the contract itself.”); see also RESTATEMENT (SECOND) OF CONTRACTS § 214, cmt. b (1981) (“Any determination of meaning or ambiguity should only be made in the light of the relevant evidence of the situation and relations of the parties, the subject matter of the transaction, preliminary negotiations . . . , usages of trade, and the course of dealing between the parties.”).

corners of the parties' written agreement.²⁷⁹ Such evidence includes prior negotiations, statements made at the time the contract was executed, the surrounding commercial circumstances (such as market conditions), course of performance, course of dealing, and usages of trade.²⁸⁰

To elaborate, textualist jurisdictions typically follow what is called the "plain meaning rule" or "four corners rule."²⁸¹ That rule sets forth a two-stage process.²⁸² During the first stage, the court assesses whether the contract is ambiguous.²⁸³ An ambiguity exists when the relevant contractual language is "reasonably susceptible" to more than one meaning.²⁸⁴ The ambiguity determination is a question of law for the judge.²⁸⁵ And in making that determination, the judge may consider only the contract itself; the investigation is restricted to the "four corners" of the document.²⁸⁶ If the

²⁷⁹ *Nautilus Marine Enterprises, Inc. v. Exxon Mobile Corp.*, 305 P.3d 309, 316 (Alaska 2013).

²⁸⁰ PERILLO, *supra* note 271, § 3.9, at 128–29. For an excellent overview of the types of extrinsic evidence, see BURTON, *supra* note 9, ch. 2, at 35–62.

²⁸¹ See MARGARET N. KNIFFIN, 5 CORBIN ON CONTRACTS § 24.7, at 33 (Joseph M. Perillo ed., rev. ed. 1998); Aaron D. Goldstein, *The Public Meaning Rule: Reconciling Meaning, Intent, and Contract Interpretation*, 53 SANTA CLARA L. REV. 73, 75 (2013). Courts often use the descriptions "four-corners rule" and "plain meaning rule" synonymously. See, e.g., *In re Zecevic*, 344 B.R. 572, 578 (Bankr. N.D. Ill. 2006); *Benz v. Town Center Land, LLC*, 314 P.3d 688, 694 (N.M. Ct. App. 2013); *Gary's Implement, Inc. v. Bridgeport Tractor Parts, Inc.*, 702 N.W.2d 355, 376 (Neb. 2005); but see BURTON, *supra* note 9, § 4.2.1, at 111, § 6.3, at 224–25 (distinguishing the "four corners rule" from the "plain meaning rule"). And sources frequently distinguish between the "plain meaning rule" and the "context rule." See, e.g., *Brown v. Scott Paper Worldwide Co.*, 20 P.3d 921, 929 (Wash. 2001); Goldstein, *supra*, at 75. But some scholars use the phrase "plain meaning rule" more broadly to refer to both textualist authorities and most contextualist authorities. See, e.g., FARNSWORTH, *supra* note 270, § 7.12, at 466; PERILLO, *supra* note 271, § 3.10, at 129–30.

²⁸² FARNSWORTH, *supra* note 270, § 7.12, at 463.

²⁸³ *Id.*

²⁸⁴ KNIFFIN, 5 CORBIN ON CONTRACTS, *supra* note 281, § 24.7, at 33–34 & 41–42 (explaining that both textualist and contextualist courts use this definition of ambiguity); see, e.g., *Pioneer Peat, Inc. v. Quality Grassing & Services, Inc.*, 653 N.W.2d 469, 473 (Minn. Ct. App. 2002) (textualist decision); *California Teachers' Ass'n v. Governing Bd. of Hilmar Unified School Dist.*, 115 Cal. Rptr. 2d 323, 328 (Cal. Ct. App. 2002) (contextualist decision).

²⁸⁵ *Gulf Ins. Co. v. Burns Motor, Inc.*, 22 S.W.3d 417, 423 (Tex. 2000); *W.W.W. Associates, Inc. v. Giancontieri*, 566 N.E.2d 639, 643 (N.Y. 1990); *Quake Construction, Inc. v. American Airlines, Inc.*, 565 N.E.2d 990, 994 (Ill. 1990); PERILLO, *supra* note 271, § 3.10, at 131.

²⁸⁶ BURTON, *supra* note 9, § 4.2.2, at 111–12; KNIFFIN, 5 CORBIN ON CONTRACTS, *supra* note 281, § 24.7, at 33. Note that in assessing ambiguity, textualist courts generally interpret the document "in light of rules of grammar and the canons of construction." BURTON, *supra*, § 4.3.2, at 126; see generally *id.* § 2.4, at 57–60 (surveying the canons of construction); FARNSWORTH, *supra* note 270, § 7.10, at 456–61 (same). They also use dictionaries. BURTON, *supra*, § 2.1.2, at 38. It is only evidence from beyond the four corners that is forbidden. *Id.* § 4.3.2, at 126.

court concludes that the contract is unambiguous, it simply applies the unambiguous, “plain meaning” of the language to the facts of the case.²⁸⁷ Extrinsic evidence is never reviewed by the judge.²⁸⁸ And the case can be disposed of via a motion to dismiss, a motion for summary judgment, or some other preliminary proceeding.²⁸⁹ If the judge concludes that the contract is ambiguous, then interpreting the agreement moves to the second stage—resolving the ambiguity. At that stage, interpretation is a question of fact²⁹⁰ and extrinsic evidence regarding the contract’s meaning is admissible.²⁹¹ Generally, ambiguities are resolved by the jury when (1) it is the finder of fact, and (2) the resolution “depends on disputed extrinsic evidence.”²⁹² The judge resolves the ambiguity in all other circumstances.²⁹³ Because textualist courts conduct the initial ambiguity determination without considering materials beyond the four corners of the document, the text of the contract is often the only evidence reviewed in ascertaining the meaning of the agreement. Hence the name of this interpretive school: “textualism.”

Contextualism generally involves the same two-stage process as textualism.²⁹⁴ But the contextualist approach differs in the method used for establishing whether a contract is ambiguous. According to this view, both the language of the agreement *and* extrinsic evidence are relevant in

²⁸⁷ BURTON, *supra* note 9, § 4.2.3, at 118 (“If the document does not appear to be ambiguous, the analysis ends; the plain meaning rule comes into play to require that the judge give the unambiguous meaning to the contract as a matter of law.”).

²⁸⁸ *Id.* (“No extrinsic evidence then is admissible for the purpose of giving meaning to the writing.”).

²⁸⁹ *Saregama India Ltd. v. Mosley*, 635 F.3d 1284, 1290 (11th Cir. 2011); *Abundance Partners LP v. Quamtel, Inc.*, 840 F. Supp. 2d 758, 767 (S.D.N.Y. 2012); *Seaco Ins. Co. v. Barbosa*, 761 N.E.2d 946, 951 (Mass. 2002).

²⁹⁰ *Seaco Inc. Co.*, 761 N.E.2d at 951.

²⁹¹ BURTON, *supra* note 9, § 4.2.3, at 118 (“If the contract is ambiguous on its face, extrinsic evidence is admissible for” the purpose of interpreting the contract.).

²⁹² *Id.*

²⁹³ *See Pamado, Inc. v. Hedinger Brands, LLC*, 785 F. Supp. 2d 698, 707–08 (N.D. Ill. 2011); BURTON, *supra* note 9, § 4.2.3, at 118.

²⁹⁴ *See FARNSWORTH, supra* note 270, § 7.12, at 466–67 (explaining that *Pacific Gas & Electric Co. v. G.W. Thomas Drayage & Rigging Co.*, 442 P.2d 641 (Cal. 1968), the foundational and seminal contextualist case, endorsed the same two-stage process used by textualist authorities); BURTON, *supra* note 9, § 4.2.2, at 112–14; *see generally id.* § 4.1, at 106–20 (outlining both the textualist and contextualist approaches to the ambiguity determination).

deciding whether an ambiguity exists.²⁹⁵ In other words, at stage one, the judge must consider extrinsic evidence proffered by the parties, something forbidden by textualism. However, the ambiguity issue is still a question of law for the judge.²⁹⁶ And it can be resolved via summary judgment or at trial by holding an evidentiary hearing or ruling upon a motion for a directed verdict.²⁹⁷

Both textualist and contextualist courts generally consider all extrinsic evidence admissible at stage two once a contract is determined to be ambiguous.²⁹⁸ The essence of their disagreement is over what evidence may be considered during stage one in making the ambiguity determination.²⁹⁹

In sum, under textualism, before extrinsic evidence of the context may be admitted, ambiguity must be apparent on the face of the agreement.³⁰⁰

²⁹⁵ BURTON, *supra* note 9, § 4.2.2, at 112.

²⁹⁶ *Id.* § 4.2.3, at 118–19.

²⁹⁷ BNC Mortgage, Inc. v. Tax Pros, Inc., 46 P.3d 812, 819–20 (Wash. Ct. App. 2002), *overruled on other grounds*, Columbia Community Bank v. Newman Park, LLC, 304 P.3d 468, 472 (Wash. 2013); BURTON, *supra* note 9, § 4.2.3, at 118–19.

²⁹⁸ BURTON, *supra* note 9, § 1.2.3, at 14 (“Under the prevailing law, all of the elements [of extrinsic evidence] are available after a court has determined that a contract is ambiguous.”); *accord id.* ch. 5, at 151; Schwartz & Scott, *supra* note 9, at 963 n.94 (“But what if there is a genuine ambiguity in the written agreement? In such a case, the divide between formalist and anti-formalist positions essentially disappears: a court will consider extrinsic evidence to resolve the ambiguity.”); *see, e.g.*, Bank of New York Trust Co., N.A. v. Franklin Advisers, Inc., 726 F.3d 269, 276 (2d Cir. 2013) (applying New York law) (textualist decision); Wagner v. Columbia Pictures Industries, Inc., 52 Cal. Rptr. 3d 898, 901 (Cal. Ct. App. 2007) (contextualist decision). Note that some scholars support using a narrower range of evidence to resolve ambiguities. *See, e.g.*, BURTON, *supra* note 9, §§ 6.1.2.2 & 6.1.3, at 211–12 (arguing that only “objective” evidence should be considered both in determining whether an ambiguity exists and in resolving ambiguities); *see also infra* notes 306–09 and accompanying text (explaining the difference between objective and subjective evidence).

²⁹⁹ *See* Goldstein, *supra* note 281, at 80 (“The various jurisdictions then diverge as to what additional evidence [beyond the language of the contract] courts should consider to determine whether the contract is ambiguous.”); BURTON, *supra* note 9, § 4.2.2, at 111 (“On the question of ambiguity, there is a significant controversy among the courts.”).

Note that there are some judicial opinions that could be read as dispensing entirely with stage one—the ambiguity determination. *See* BURTON, *supra* note 9, § 4.3.3.2, at 131–34 (collecting authorities). But I think most such decisions are better read as standard examples of contextualism that preserve stage one. *See also id.* § 6.1.2.1, at 204–05 (arguing that the procedural posture of a contract interpretation case requires the court to make an ambiguity determination because, under the rules of civil procedure, if the contract is unambiguous, the court must decide the case as a matter of law).

³⁰⁰ BURTON, *supra* note 9, § 4.2.2, at 112; *see, e.g.*, IDT Corp. v. Tyco Grp., 918 N.E.2d 913, 916 (N.Y. 2009).

Such an ambiguity is typically called “patent,” “intrinsic,” or “facial.”³⁰¹ Under contextualism, extrinsic evidence of the context may be used to establish that facially clear language is actually ambiguous.³⁰² Such an ambiguity is typically called “latent” or “extrinsic.”³⁰³ Put simply, textualism recognizes only patent ambiguities, whereas contextualism recognizes both patent and latent ambiguities.

While most scholars and many courts endorse the basic framework set forth above,³⁰⁴ it must be noted that this framework is a considerable oversimplification of the caselaw.³⁰⁵ For example, contextualist courts can be subdivided based on the type of extrinsic evidence they permit in assessing contractual ambiguity. Some courts consider all relevant evidence in making the ambiguity determination.³⁰⁶ Others limit the inquiry to “objective” evidence—i.e., evidence that can be provided by disinterested third parties, such as testimony regarding usages of trade.³⁰⁷ “Subjective evidence,” such as testimony from the parties about what they believed the contract meant and preliminary negotiations, may not be used to establish an ambiguity.³⁰⁸ This distinction is often justified on the ground that objective evidence is more reliable than subjective evidence because it is much more difficult to fabricate.³⁰⁹ Courts following the plain meaning rule

³⁰¹ See *Watkins v. Ford*, 304 P.3d 841, 847 (Utah 2013); BURTON, *supra* note 9, § 4.1, at 107; FARNSWORTH, *supra* note 270, § 7.12, at 464.

³⁰² BURTON, *supra* note 9, § 4.2.2, at 112; see, e.g., *Shay v. Aldrich*, 790 N.W.2d 629, 641 (Mich. 2010).

³⁰³ See BURTON, *supra* note 9, § 4.1, at 107; FARNSWORTH, *supra* note 270, § 7.12, at 464 & n.16.

³⁰⁴ For several examples, see *supra* note 274; but see Margaret N. Kniffin, *Conflating and Confusing Contract Interpretation and the Parol Evidence Rule: Is the Emperor Wearing Someone Else's Clothes?*, 62 RUTGERS L. REV. 75, 95 (2009) (dividing the cases into three broad schools rather than two).

³⁰⁵ GERARD MCMEEL, *THE CONSTRUCTION OF CONTRACTS: INTERPRETATION, IMPLICATION, AND RECTIFICATION* § 1.31 (2d ed. 2011) (explaining that dividing the interpretation caselaw into literalist and purposivist schools is “too simplistic”).

³⁰⁶ See, e.g., *Ward v. Intermountain Farmers Ass'n*, 907 P.2d 264, 268 (Utah 1995); *Adams v. MHC Colony Park Limited P'ship*, 169 Cal. Rptr. 3d 146, 161 (Cal. Ct. App. 2014); see also BURTON, *supra* note 9, § 4.2.2, at 112–14 & 117 (explaining this approach).

³⁰⁷ See, e.g., *AM Int'l, Inc. v. Graphic Mgmt. Assoc., Inc.*, 44 F.3d 572, 575 (7th Cir. 1995); see also BURTON, *supra* note 9, § 4.2.2, at 114–15 & 117 (explaining this approach).

³⁰⁸ See *AM Int'l*, 44 F.3d at 575; BURTON, *supra* note 9, § 4.2.2, at 115.

³⁰⁹ See, e.g., *AM Int'l*, 44 F.3d at 575; PERILLO, *supra* note 271, § 3.10, at 130.

can also be subdivided in various ways.³¹⁰ In reality, there are “innumerable gradations” of textualism and contextualism in the caselaw.³¹¹

Moreover, the law of contract interpretation is extraordinarily convoluted. “In virtually every jurisdiction, one finds irreconcilable cases, frequent changes in doctrine, confusion, and cries of despair.”³¹² The precise formulation of a rule is frequently inconsistent with the way the rule is applied.³¹³ And courts sometimes set forth inconsistent rules within a single opinion.³¹⁴ As a result, most states fall somewhere between

³¹⁰ See Goldstein, *supra* note 281, at 75 n.2 (“There are also variants of the plain meaning rule that differ in the strictness with which courts limit themselves to the text of the contract alone.”); PETER LINZER, 6 CORBIN ON CONTRACTS § 25.13, at 146 (Joseph M. Perillo ed., rev. ed. 2010) (explaining that some textualist courts take a narrow view of what constitutes a facial ambiguity, while others have “a greater willingness” to find such ambiguities).

³¹¹ LINZER, 6 CORBIN ON CONTRACTS, *supra* note 310, § 25.13, at 146; see also Peter Linzer, *The Comfort of Certainty: Plain Meaning and the Parol Evidence Rule*, 71 FORDHAM L. REV. 799, 805–06 (2002) (“A detailed survey will reveal countless variations around the country and remarkable gradations of what seem to be fixed rules, even within a given jurisdiction.”).

³¹² Eric A. Posner, *The Parol Evidence Rule, The Plain Meaning Rule, and the Principles of Contract Interpretation*, 146 U. PA. L. REV. 533, 540 (1998); accord PERILLO, *supra* note 271, § 3.1, at 106 (noting that the courts do not consistently follow the rules of contract interpretation); *id.* § 3.2(b), at 110 n.29 (collecting secondary authorities that address the confused state of the law in Alaska, California, Illinois, Montana, Oregon, Texas, and Wisconsin and further noting that “[o]ther jurisdictions could be cited”); LORD, *supra* note 271, § 33:42, at 1190 (“Not only do various jurisdictions disagree as to how and when extrinsic evidence of the circumstances surrounding the execution of a contract becomes admissible, but the decisions within a given jurisdiction are often difficult and sometimes impossible to reconcile on this point.”). For my favorite “cry of despair,” see *Jake C. Byers, Inc. v. J.B.C. Investments*, 834 S.W.2d 806, 811 (Mo. App. 1992).

³¹³ See LINZER, 6 CORBIN ON CONTRACTS, *supra* note 310, § 25.14[A], at 148–61 (collecting examples).

³¹⁴ *Id.* § 25.15[c], at 192 (“At times a state court seems to be saying contradictory things.”); *id.* at 195 (discussing as an example *Wadi Petroleum v. Ultra Resources*, 65 P.3d 703, 706–10 (Wy. 2003)). Consider the *Restatement (Second) of Contracts*. Some commentators believe the *Restatement* endorses a standard version of contextualism. See, e.g., PERILLO, *supra* note 271, § 3.12, at 135. Others contend that the *Restatement* goes further and dispenses with the ambiguity determination entirely. See, e.g., BURTON, *supra* note 9, §§ 4.2.2, at 115–17, & 4.5.2. This confusion is understandable, for the *Restatement* contains language strongly supporting both positions. In two places, the *Restatement* provides that extrinsic evidence is only admissible to support an interpretation if the language of the parties’ contract is “reasonably susceptible” to the proffered reading. See RESTATEMENT (SECOND) OF CONTRACTS § 213, cmt. b (1981); *id.* § 215, cmt. b. These provisions require that courts conduct a standard ambiguity inquiry. However, the *Restatement* also says that words can mean whatever the parties intend them to mean. See *id.* § 201(1) (“Where the parties have attached the same meaning to a promise or agreement or a term thereof, it is interpreted in accordance with that meaning.”); *id.* § 212 cmt. b, illus. 3 & 4 (each providing that parties can use words to mean the exact opposite of their standard meaning). These provisions indicate that the ambiguity analysis is unnecessary; a proffered interpretation supported by extrinsic evidence need not fit the language of the parties’ contract at all, at

textualism and contextualism, rather than firmly on one side.³¹⁵ This likely explains why commentators differ over which is the majority approach in this country.³¹⁶

least as such language is generally understood. *See also* Helen Hadjiyannakis, *The Parol Evidence Rule and Implied Terms: The Sounds of Silence*, 54 *FORDHAM L. REV.* 35, 60 n.134 (1985) (explaining that the portions of § 212 cited immediately above endorse the view that parties may adopt a “private code” to be used in interpreting their agreement, under which words can mean the exact opposite of their meaning under standard usage); KNIFFIN, 5 *CORBIN ON CONTRACTS*, *supra* note 281, § 24.8, at 54–59 (discussing private codes generally). If the drafters of the *Restatement* were this confused, it should not be surprising that generations of courts, lawyers, and law students have struggled with the principles of contract interpretation and the parol evidence rule.

³¹⁵ There are two generally prevailing theories as to why this confusion exists. Some believe that it is because courts fail to carefully distinguish between interpretive principles like the plain meaning rule, on the one hand, and the parol evidence rule, on the other hand. *See, e.g.*, BURTON, *supra* note 9, §§ 3.1, at 64, & 4.2.4, at 120; *see also* Kniffin, *supra* note 304 (discussing how courts and scholars confuse interpretation and the parol evidence rule and the injustice that results). Others suggest that it is because interpretation and the parol evidence rule are exceptionally difficult to distinguish. *See, e.g.*, PERILLO, *supra* note 271, § 3.9, at 201; Linzer, *supra* note 311, at 801; *see also* FARNSWORTH, *supra* note 270, § 7.12, at 466 (“The question then is: where does ‘interpretation’ end and ‘contradiction’ or ‘addition’ begin?”). I think both explanations have considerable validity.

³¹⁶ Compare BURTON, *supra* note 9, § 4.3.2, at 126 (“Most courts follow the four corners rule when deciding whether a contract is ambiguous, sometimes under the guise of the parol evidence rule.”), and Schwartz & Scott, *supra* note 9, at 928 n.1 (“A strong majority of U.S. courts continue to follow the traditional, ‘formalist’ approach to contract interpretation. A state-by-state survey of recent court decisions shows that thirty-eight states follow the textualist approach to interpretation. Nine states, joined by the Uniform Commercial Code for sales cases (U.C.C.) and the Restatement (Second) of Contracts, have adopted a contextualist or ‘antiformalist’ interpretive regime. The remaining states are indeterminate.”), with LORD, *supra* note 271, § 30:5, at 80 (“While there is authority that the court is limited in its consideration solely to the face of the written agreement, many more courts take the position that a court may provisionally receive all credible evidence concerning the parties’ intentions to determine whether the language of the contract is reasonably susceptible to the interpretation urged by the party claiming ambiguity; if it is, this evidence may then be admitted and heard by the trier of fact.”). Note also that in cases involving the sale of goods, the Uniform Commercial Code expressly requires that courts use a form of contextualist interpretation. *See* U.C.C. § 1-303(d); *id.* § 2-202 & cmt. 1(c). But at least one respected source contends that judges frequently ignore the dictates of Article 2 and use a form of textualism, requiring a finding of facial ambiguity before admitting extrinsic evidence for interpretive purposes. *See* 1 JAMES J. WHITE ET AL., *UNIFORM COMMERCIAL CODE* § 3:14, at 221–22 (6th ed. 2010).

The picture appears to be clearer abroad, with contextualism now dominate both in other nations and in international law. *See* MCMEELE, *supra* note 305, § 2.01 (explaining that the general trend in common-law jurisdictions is towards adoption of the contextualist approach); CATHERINE MITCHELL, *INTERPRETATION OF CONTRACTS: CURRENT CONTROVERSIES IN LAW* 58 (2007) (explaining that the same trend exists in European civil-law jurisdictions); United Nations Convention on Contracts for the International Sales of Goods, art. 8(3), Apr. 11, 1980, *available at* <http://www.uncitral.org/pdf/english/texts/sales/cisg/V1056997-CISG-e-book.pdf>; *Unidroit Principles of International Commercial Contracts*, art. 4.3 (2010), *available at* <http://www.unidroit.org/english/principles/contracts/principles2010/>

Like the courts, contracts scholars can also generally be divided into textualist and contextualist camps,³¹⁷ with a clear majority falling into the latter group.³¹⁸ But commentators have also proposed positions that do not fit precisely into the textualist-contextualist continuum. For example, some believe that different interpretive approaches should be applied to different types of contracts, often distinguishing commercial agreements between businesses from consumer and employment agreements.³¹⁹ Others argue that the parties should be permitted to decide which interpretive approach will be used should a dispute over contractual meaning arise.³²⁰ Some of these positions are discussed further in the next part.

B. *The Policy Debate*

Contract interpretation is an exceptionally important subject.³²¹ And the central policy issue in this field is the proper role of extrinsic evidence in the interpretive process.³²² The rules regarding such evidence can

integralversionprinciples2010-e.pdf; Principles of European Contract Law, art. 5:102 (1998), available at <http://www.jus.uio.no/lm/eu.contract.principles.parts.1.to.3.2002/>.

³¹⁷ Juliet P. Kostritsky, *Plain Meaning vs. Broad Interpretation: How the Risk of Opportunism Defeats a Unitary Default Rule for Interpretation*, 96 KY. L.J. 43, 43–44 (2007–2008).

³¹⁸ Schwartz & Scott, *supra* note 9, at 938 (“The (almost) scholarly consensus shares the UCC and Restatement view . . . that [courts should be permitted] to access a broad evidentiary base in determining both the terms of the contract and the meaning to be attached to those terms.”); accord PERILLO, *supra* note 271, § 3.10, at 130. For older sources supporting a contextualist approach to interpretation, see, for example, JAMES BRADLEY THAYER, A PRELIMINARY TREATISE ON THE LAW OF EVIDENCE 428–29 (1898); 9 JOHN HENRY WIGMORE, A TREATISE ON THE ANGLO-AMERICAN SYSTEM OF EVIDENCE IN TRIALS AT COMMON LAW § 2470 (3d ed. 1940); and Arthur L. Corbin, *The Interpretation of Words and the Parol Evidence Rule*, 50 CORNELL L.Q. 161, 188–89 (1965). For modern sources, see, for example, BURTON, *supra* note 9, § 6.1.2.1, at 209; PERILLO, *supra*, § 3.10, at 130; FARNSWORTH, *supra* note 270, § 7.10, at 453–54; KNIFFIN, 5 CORBIN ON CONTRACTS, *supra* note 281, § 24.7, at 36–39; and LINZER, 6 CORBIN ON CONTRACTS, *supra* note 310, §§ 25.14, at 39, & 25.14[B], at 163.

³¹⁹ See *infra* notes 425–28 and accompanying text.

³²⁰ See *infra* notes 429–36 and accompanying text.

³²¹ See *supra* notes 9–11 and accompanying text.

³²² See Hermalin et al., *supra* note 9, at 88–89 (“The key policy question underlying contract interpretation is how thorough the interpretive process should be; and this question is commonly articulated in terms of the dichotomy of form and substance.”); see also PETER A. ALCES, A THEORY OF CONTRACT LAW: EMPIRICAL INSIGHTS AND MORAL PSYCHOLOGY 153 (2011) (“The parol or extrinsic evidence tension in contract is fundamental; it concerns the very foundations of agreement”); MCMEELE, *supra* note 305, § 5.01 (“One of the most controversial areas in the principles governing the interpretation of contracts is the question of what materials are admissible to assist the court in carrying

influence virtually every aspect of the parties' contractual relationship.³²³ Thus, numerous factors are potentially relevant in deciding the optimal interpretive regime.³²⁴ Given these points, it should not be surprising that the debate among courts and scholars over the proper approach to extrinsic evidence has been fierce,³²⁵ with some judges adopting "sky-is-falling" rhetoric when criticizing the opposition.³²⁶ The disputants have advanced a wide array of conceptual, theoretical, and empirical arguments in support of textualism and contextualism.³²⁷ But the debate has focused on three basic topics: (1) interpretive accuracy; (2) transaction costs; and (3) enforcement costs.³²⁸

out the task.") (focusing on contract interpretation in jurisdictions outside the United States, particularly England and other common law nations); Goldstein, *supra* note 281, at 74 ("When and what kinds of extrinsic evidence should courts admit in order to interpret the meaning of a contract? . . . [T]he answer has profound implications for whether courts achieve the goals of predictability and fairness that motivate the law of contracts."); Steven Shavell, *On the Writing and Interpretation of Contracts*, 22 J.L. ECON. & ORG. 289, 311 (2006) ("Third, the issue of the use of evidence extrinsic to contracts in their interpretation is of significance in actual practice and is also much debated. . . . [T]he question whether the value of extrinsic evidence in contractual interpretation exceeds its costs to the parties is a real one.").

³²³ Hermalin et al., *supra* note 9, at 90 ("The regime of contract interpretation will influence the contracting parties' behavior in many respects: with regard to decisions to breach, to take advance precautions, to mitigate damages, to gather and communicate information, to allocate risk, to make reliance investments, to behave opportunistically, and to spend resources in litigation, and so on.").

³²⁴ *Id.* ("The considerations that determine the optimal approach to contract interpretation are thus quite broad-ranging."); *see id.* at 90–91 (setting forth a list of some of the key considerations).

³²⁵ Kostritsky, *supra* note 317, at 54 (concerning scholars); McLauchlan, *supra* note 10, at 5 ("There are fundamental divisions among commentators, practitioners and judges . . . as to the nature of the task and the permissible aids to interpretation.").

³²⁶ For perhaps the best example, see *infra* notes 395–98 and accompanying text.

³²⁷ *See* Burton, *supra* note 10, at 341 ("Among contract scholars, there is no consensus about how an interpreter should accomplish these tasks. Consequently, normative theories of contract interpretation proliferate."). For an outstanding overview of the economic arguments for and against the two approaches to contract interpretation, see Cohen, *supra* note 272.

³²⁸ *See* Schwartz & Scott, *supra* note 9, at 930 ("Since no interpretive theory can justify devoting infinite resources to achieving interpretive accuracy, any socially-desirable interpretive rule would trade off accuracy against contract-writing and adjudication costs."); Cohen, *supra* note 272, at 148 ("The real question is which [interpretive] methodology has the lowest error rate and at what cost."); J.J. Spigelman, *Contractual Interpretation: A Comparative Perspective*, 85 AUSTR. L.J. 412, 413 (2011) ("Like many other aspects of contract law, interpretation requires the resolution of a tension between certainty or efficiency on the one hand and accuracy or fairness on the other."); *see also* MITCHELL, *supra* note 316, at 108 ("The first and most obvious reason for confining a court's enquiry to the four corners of the agreement relates to the possible costs involved in the contextual approach. . . . Two particular kinds of costs are pertinent: *transaction costs* (broadly, the cost of reaching and recording the

1. *Interpretive Accuracy*

As noted above, the purpose of contract interpretation is to ascertain the intent of the parties.³²⁹ But which school of interpretation best accomplishes this task? Both sides maintain it is their approach.³³⁰ Adherents of textualism typically offer the following arguments in support of their position. First, the express terms of a contract are the best evidence of contractual intent.³³¹ The parties likely chose the words of their agreement with care to reflect their mutual understanding.³³² By contrast, contextual evidence is often unreliable, particularly in light of the problems with memory.³³³ And such evidence is frequently ambiguous and/or contradictory.³³⁴ Second, judges can more skillfully apply textualist methodology. They are better at adopting the perspective of “a reasonable recipient of the document” who is focused on the contractual language than

deal) and *enforcement costs* (broadly, the cost of insuring compliance in resolving disputes.)” (emphasis added)).

³²⁹ See *supra* note 271 and accompanying text.

³³⁰ See Cohen, *supra* note 272, at 145 (“Scholars disagree, however, over whether strict approaches to interpretation and implied terms, such as textualism, lead to more court error than broader approaches, such as contextualism.”); see also *Steuart v. McChesney*, 444 A.2d 659, 662 (Pa. 1982) (noting that textualism “has been supported as generally best serving the ascertainment of the contracting parties’ mutual intent”).

³³¹ *ALCES*, *supra* note 322, at 149 (“The case favoring ‘plain meaning’ is clear: Courts cannot read the minds of litigants, so the clear expression of their intent is the best evidence of what that intent actually is (or, at least, was).”); Cohen, *supra* note 272, at 131 (explaining that textualists presume that “the express terms of the contract . . . best approximate the parties’ intentions”); see also *MITCHELL*, *supra* note 316, at 94 (“Rather, the hallmark of a more serious kind of formalism in contract would be the tendency to regard the contractual text as supreme evidence of the parties’ intentions, over more elusive and equivocal evidential material, such as trade customs, previous dealings and so on”); *Slamow v. Del Col*, 594 N.E.2d 918, 919 (N.Y. 1992) (“The best evidence of what parties to a written agreement intend is what they say in their writing.”).

³³² See *Steuart v. McChesney*, 444 A.2d 659, 662 (Pa. 1982) (“Where the contract evidences care in its preparation, it will be presumed that its words were employed deliberately and with intention. . . . Courts in interpreting a contract do not assume that its language was chosen carelessly.”) (internal quotation marks omitted).

³³³ See *Shavell*, *supra* note 322, at 311 (observing that “extrinsic evidence . . . is highly imperfect”); *Goldstein*, *supra* note 281, at 75–76 (“While the context rule responds to the issues associated with interpreting language in a vacuum, it relies upon unreliable evidence in order to give meaning to contract language. Parties lie and misremember, especially regarding extrinsic evidence such as prior negotiations, course of performance, and course of dealing.”).

³³⁴ *Goldstein*, *supra* note 281, at 76 (“Also, extrinsic evidence of parties’ prior acts is often compatible with numerous contradictory accounts of what the parties intended, and thus fails to shed light on the parties’ actual bargain.”).

the perspective of someone who participated in the preliminary negotiations and is familiar with the surrounding circumstances.³³⁵ In part, that is because judges often misunderstand the economic context in which business contracts are made, even when presented with significant evidence on the subject.³³⁶ Such evidence overwhelms them with information that they lack the training to fully comprehend.³³⁷ Moreover, the difficulties judges face in applying contextualist methodology are particularly acute when the parties assert that they were using some type of private language (such as an industry dialect) in which words have a meaning that differs from standard usage.³³⁸ Third, the quantity and quality of evidence available under contextualism confuses juries, leading to more mistakes in ascertaining contractual meaning.³³⁹ Fourth, the first three arguments are strengthened by the fact that parties often intentionally use extrinsic evidence to distort the meaning of an otherwise clear agreement. Under contextualism, parties have an incentive to sift through the preliminary negotiations, prior dealings, and industry custom, in the search for some remark or practice that can be deployed to alter the contract and obtain an unbargained-for advantage.³⁴⁰ And judges have difficulty policing such manipulated contextual evidence.³⁴¹

³³⁵ MITCHELL, *supra* note 316, at 116 (explaining this position).

³³⁶ See VICTOR GOLDBERG, *FRAMING CONTRACT LAW: AN ECONOMIC PERSPECTIVE* 163 (2006); see also MITCHELL, *supra* note 316, at 116 (“A court’s conclusions on the social context of a commercial agreement may be impressionistic at best, despite hearing testimony of witnesses and experts.”).

³³⁷ See MITCHELL, *supra* note 316, at 115 (“The contextual approach arguably increases the chances for error by increasing the amount of information deemed relevant to the interpretation exercise. Judges may have to deal with a significant amount of contextual material, some of it connected to particular frameworks of analysis whose conventions will be unfamiliar to them.”).

³³⁸ See Alan Schwartz & Robert E. Scott, *Contract Theory and the Limits of Contract Law*, 113 *YALE L.J.* 541, 587 (2003).

³³⁹ See KNIFFIN, 5 *CORBIN ON CONTRACTS*, *supra* note 281, § 24.7, at 53 (noting that judges often argue that “juries might incorrectly assess the extrinsic evidence” that is admissible under a contextualist approach); see, e.g., *W.W.W. Assoc., Inc. v. Giancontieri*, 566 N.E.2d 639, 642 (N.Y. 1990); see also MITCHELL, *supra* note 316, at 110 (“[T]he greater the amount of contextual material, the greater the possibility for error. Decision-makers may easily become bewildered by a large set of conflicting evidence.”) (internal quotation marks omitted); LINZER, 6 *CORBIN ON CONTRACTS*, *supra* note 310, § 25.17, at 241 & n.4 (“It is a commonplace that underlying the restrictive use of the parol evidence rule is distrust of juries in contract cases.”).

³⁴⁰ MITCHELL, *supra* note 316, at 113 (“A further problem is that much reliance on context may be done strategically—the problem of ‘threshing through the undergrowth’ for the chance remark upon which to build a case. The suspicion is often raised of the strategic reliance on context to sanction an

Contextualists' primary argument on the issue of interpretive accuracy is that meaning can only be determined by considering the context in which language is used.³⁴² First, as dictionaries demonstrate, "most words have several meanings in the abstract (acontextually)."³⁴³ Second, words can have meanings that are not set forth in dictionaries, meanings that textualist interpretation largely ignores.³⁴⁴ Third, it is sometimes impossible to perceive an ambiguity without reviewing evidence from beyond the four corners of the contract.³⁴⁵ For example, in the famous case of *Raffles v. Wichelhaus*,³⁴⁶ the parties' agreement provided, in perfectly clear terms, that certain cotton would arrive on the ship *Peerless*. But there were two ships with that name, creating an ambiguity that became apparent only upon the consideration of extrinsic evidence.³⁴⁷ Judge Posner explains this problem as follows: "The contract's words point out to the real world, and the real world may contain features that make seemingly clear words,

escape from a bad bargain[.] . . . even in circumstances where the written terms appear relatively complete."); *accord* *Steuart v. McChesney*, 444 A.2d 659, 663 (Pa. 1982) ("Likewise, resort to the plain meaning of language hinders parties dissatisfied with their agreement from creating a myth as to the true meaning of the agreement through subsequently exposed extrinsic evidence."); Linzer, *supra* note 311, at 804 (offering the following as a possible justification for the plain meaning rule: "strict rules protect against the fear that the more we allow the words of a contract to be challenged in the name of the parties' actual intent, the more we produce disorder and even chaos, waiting to be exploited by unscrupulous litigants who demand a bonus to do what they already promised to do").

³⁴¹ See Cohen, *supra* note 272, at 145–46.

³⁴² See, e.g., RESTATEMENT (SECOND) OF CONTRACTS § 202 cmt. d (1981) ("Meaning is inevitably dependent on context."); FARNSWORTH, *supra* note 270, § 7.10, at 454 ("Indeed, it is questionable whether a word has a meaning at all when divorced from the circumstances in which it is used."); KNIFFIN, 5 CORBIN ON CONTRACTS, *supra* note 281, § 24.7, at 36 ("Before the meaning of words and a contract can be plain and clear, at least some of the surrounding circumstances must be known[.]").

³⁴³ BURTON, *supra* note 9, § 6.1.2.2, at 210.

³⁴⁴ See Goldstein, *supra* note 281, at 75 ("The plain meaning rule also ties the interpretation of contract terms to a judge's subjective notions of what words mean in language and prevents parties from submitting evidence of alternate meanings that may be publically used and acknowledged, but not set forth in a standard dictionary.").

³⁴⁵ RESTATEMENT (SECOND) CONTRACTS § 214, cmt. b (1981) ("Even though words seem on their face to have only a single possible meaning, other meanings often appear when the circumstances are disclosed."); FARNSWORTH, *supra* note 270, § 7.12, at 465 (same).

³⁴⁶ 159 Eng. Rep. 375 (Exch. 1864).

³⁴⁷ *Charter Oil Co. v. American Employers' Ins. Co.*, 69 F.3d 1160, 1167 (D.C. Cir. 1995); LINZER, 6 CORBIN ON CONTRACTS, *supra* note 310, § 25.4, at 32.

sentences, and even entire documents ambiguous.”³⁴⁸ Given these points, textualism is impossible; “plain meaning” simply does not exist.³⁴⁹ Excluding evidence from outside the four corners thus fatally undermines the interpretive process.³⁵⁰ And this position finds support in the philosophy of language, linguistics, and other related fields.³⁵¹

Similarly, contextualists argue that it is impossible to exclude material from outside the four corners of the contract when engaging in interpretation. The only question is what type of extrinsic evidence gets emphasized—the background of the judge (under textualism) or the background of the parties to the transaction (under contextualism).³⁵² The latter, contextualists argue, is clearly more useful in attempting to ascertain the intent of the parties.³⁵³ For example, judges frequently come from backgrounds that are quite different from the “the specialized worlds of trade” that serve as the context for many commercial agreements.³⁵⁴ In such

³⁴⁸ Posner, *supra* note 195, at 1597 (then offering *Raffles* as an example); *accord* Charter Oil Co., 69 F.3d at 1167 (“Latent ambiguity can arise where language, clear on its face, fails to resolve an uncertainty when juxtaposed with circumstances in the world that the language is supposed to govern.”); *see also* KNIFFIN, 5 CORBIN ON CONTRACTS, *supra* note 281, § 24.7, at 36 (“[P]roof of the circumstances may make plain and clear a meaning that was not apparent when in the absence of such proof some other meanings seemed plain and clear.”).

³⁴⁹ *See* BURTON, *supra* note 9, § 4.6.1, at 144 (“The chief criticism of the plain meaning and four corners rules has been that there are no plain meanings that an interpreter can find on a contract document’s face.”); KNIFFIN, 5 CORBIN ON CONTRACTS, *supra* note 281, § 24.7, at 54 (“[T]he conclusions are inescapable that words used in a contract do not have only one true meaning and that words are never so ‘plain and clear’ that proof of surrounding circumstances and other extrinsic aids to interpretation can be excluded.”).

³⁵⁰ *See, e.g.*, BURTON, *supra* note 9, § 6.1.2.2, at 211 (“The necessity of context for ascertaining the meaning(s) is the strongest argument against the four corners rule here.”); PERILLO, *supra* note 271, § 3.10, at 130 (arguing that the “plain meaning rule has been properly condemned because the meaning of words varies” with changes in context); KNIFFIN, 5 CORBIN ON CONTRACTS, *supra* note 281, § 24.7, at 37 (“The plain meaning rule can exclude proof of [the parties’] actual intentions.”).

³⁵¹ *See, e.g.*, MCMEEEL, *supra* note 305, § 2.43 (“However there are lessons to be learned from these other fields. One insight stands out: in ascertaining the meaning of an utterance, the context in which it is made is indispensable.”); *see generally id.* §§ 2.37, 2.39, 2.40, 2.43, 2.57 (discussing the relevancy of the philosophy of language, linguistics, and related fields for contract interpretation).

³⁵² For the classic statement of this position, see Corbin, *supra* note 318, at 164 (“[W]hen a judge refuses to consider relevant extrinsic evidence on the ground that the meaning of written words is to him plain and clear, his decision is formed by and wholly based upon the completely extrinsic evidence of his own personal education and experience.”).

³⁵³ *See* Posner, *supra* note 312, at 560 (“The parol evidence rule excludes extrinsic evidence from consideration, while allowing the judge to rely on his or her personal knowledge, even though the former, more so than the latter, would enable the court to determine the parties’ intentions.”).

³⁵⁴ *See* Mellon Bank, N.A. v. Aetna Business Credit, 619 F.2d 1001, 1011 n.12 (3d Cir. 1980).

cases, “the parties’ linguistic reference” is far more likely to provide insight into their intent than the judge’s language background.³⁵⁵ “Using the interpreter’s context injects arbitrariness into the process; it bears no reliable relation to the parties’ intention and, indeed, may be quite foreign to them.”³⁵⁶ And evidence of the context can be particularly useful to less experienced judges, bringing their “information sets” more closely into alignment with those of experienced judges who may already have some understanding of the relevant commercial practices.³⁵⁷

Lastly, contextualists have no sympathy for textualist concerns that extrinsic evidence may confuse or fool juries. First, judges can use preliminary proceedings such as summary judgment to prevent baseless arguments from reaching the jury.³⁵⁸ Remember that even under contextualism the language of the agreement places a critical limit on the spectrum of possible interpretations: the words must be reasonably susceptible to the meanings proffered by the parties. Any reading that fails this test is barred from evidence.³⁵⁹ Second, as explained by Peter Linzer, textualists’ worries regarding juries prove too much:

If the jury system is so defective that juries cannot be allowed to hear the story of black meaning white, they should not be allowed to decide wrongful death actions, complex anti-trust suits and patent cases, much less psychological defenses in capital murder cases. Since we are not about to abolish the jury system generally, there is no reason to constrict it in the one area of interpretation of integrated agreements.³⁶⁰

And the same argument applies to the claim that judges lack the capacity to apply contextualist methodologies. In sum, textualism violates “the basic precept that a court should make its decisions based on full information, not conjecture.”³⁶¹

³⁵⁵ *See id.*

³⁵⁶ BURTON, *supra* note 9, § 2.1.3, at 40.

³⁵⁷ Avery Wiener Katz, *The Economics of Form and Substance in Contract Interpretation*, 104 COLUM. L. REV. 496, 526 (2004).

³⁵⁸ LINZER, 6 CORBIN ON CONTRACTS, *supra* note 310, § 25.14, at 163.

³⁵⁹ *Id.* § 25.16[B] at 219.

³⁶⁰ *Id.* § 25.4 at 36; *see also* Posner, *supra* note 312, at 567 (“The concern is that if juries considered all of the extrinsic evidence, rather than just the writing, they would not render good judgments. . . . If juries are incompetent, why would limiting them to certain kinds of evidence lead to a more accurate result?”).

³⁶¹ LINZER, 6 CORBIN ON CONTRACTS, *supra* note 310, § 25.14, at 163.

Textualists can respond to some of these arguments as follows. Plain meaning is not an impossibility; indeed, most meaning is plain.³⁶² And the philosophy of language does not suggest otherwise.³⁶³ That is because the “plain meaning” sought by textualism is *not* acontextual.³⁶⁴ In other words, textualists do not advocate acontextual interpretation. Rather, they simply “disagree over what is the correct context, over how much context is relevant or necessary to assessing meaning.”³⁶⁵ And, they contend, a “minimum evidentiary basis ordinarily will convey sufficient contextual information.”³⁶⁶

2. *Transaction Costs*

One of the chief arguments in favor of contextualist interpretation is that it reduces transaction costs.³⁶⁷ It does so by enabling the parties to draft less complete contracts.³⁶⁸ “Parties can write a simpler document, leaving it to the courts to fill the gaps through the process of contextual interpretation.”³⁶⁹ This saves on negotiating and drafting costs since the parties need not “reduce all the terms and standards that govern the agreement to writing.”³⁷⁰

³⁶² Ricks, *supra* note 271, at 769 (“Plain meaning rests instead on our unreflective, public, conventional practice of language use. Most meaning is plain. Part III explains that, though plain meaning is immune from attack on grounds of impossibility, . . .”); *accord* Kent Greenawalt, *A Pluralist Approach to Interpretation: Wills and Contracts*, 42 SAN DIEGO L. REV. 553, 592 (2005).

³⁶³ Ricks, *supra* note 271, *passim*.

³⁶⁴ *See* Greenawalt, *supra* note 362, at 592 (further noting that acontextual interpretation may indeed be philosophically incoherent); Ricks, *supra* note 271, at 769.

³⁶⁵ MITCHELL, *supra* note 316, at 9; *accord* Posner, *supra* note 195, at 1598.

³⁶⁶ Schwartz & Scott, *supra* note 9, at 952; *accord* MITCHELL, *supra* note 316, at 9.

³⁶⁷ Cohen, *supra* note 272, at 132 (calling this a “key economic argument for an expansive court role in interpreting and implying terms”); *see also* MITCHELL, *supra* note 316, at 108–09 (“One of the arguments in favor of contextualism over literalism is that it lowers transaction costs . . .”).

³⁶⁸ Cohen, *supra* note 272, at 132 (explaining that contextualism “enables and encourages parties to write less complete contracts than they otherwise would”).

³⁶⁹ MITCHELL, *supra* note 316, at 109; *see also* Posner, *supra* note 195, at 1600 (“Were evidence of trade usage barred in contract litigation, parties to contracts would be driven to include additional detail in their contracts . . .”).

³⁷⁰ MITCHELL, *supra* note 316, at 109; *accord* Cohen, *supra* note 272, at 132 (“Writing less complete contracts saves on drafting and negotiating costs so long as the court-supplied interpretations and terms sufficiently approximate the parties’ intentions.”).

Textualists essentially concede that contextualism reduces transaction costs. But they argue it does so at the price of greater enforcement costs.³⁷¹ That is, in part, because contextualist interpretation incentivizes parties to draft worse contracts—contracts with more open terms and ambiguities.³⁷² Under textualism, parties are motivated to spend additional time negotiating and writing their agreements, leading to better contracts that reduce litigation.³⁷³

The next section reviews the debate over which interpretive approach minimizes enforcement costs. For now, assume that textualists are correct that there is a trade-off between transaction and enforcement costs. “In balancing contracting and litigation costs, it is important to keep in mind that contracting costs are certain and incurred across all contracts, while litigation costs, though often much larger than contracting costs, are incurred in only a small fraction of contracts.”³⁷⁴ A critical question, then, is what will be less expensive: drafting a more complete contract for every transaction, a contract that can be interpreted with minimal contextual evidence when a dispute arises; or drafting shorter, less complete agreements that increase the likelihood of a lawsuit and require the assessment of substantially more material should there be litigation. Given how rare contract disputes are relative to the total number of agreements executed, one could plausibly argue that reducing transaction costs via the

³⁷¹ See Cohen, *supra* note 272, at 133–34 (explaining a model developed by Judge Posner under which, “as parties spend less on ex ante contracting and rely more on extrinsic evidence to prove their intent, drafting costs go down, but expected litigation costs rise,” because there is an increased likelihood of litigation and the expense of any litigation that does occur will be greater).

³⁷² In a somewhat different context, Cohen offers the following: “If a court is willing to ‘insure’ parties through flexible interpretation and implied terms it creates a classic moral hazard problem: the parties have less incentive to write good contracts themselves.” Cohen, *supra* note 272, at 137.

³⁷³ *Id.* at 133–34 (again explaining Posner’s model). There are reasons to be skeptical of this argument. The odds of there being a dispute—let alone a lawsuit—over any given contract are incredibly low. See BURTON, *supra* note 9, § 1.1.2, at 12 (“And, in light of the millions of contracts concluded each day, interpretive disputes must be rare; by far, most contracts are performed without a hitch.”). Thus, the incentives created by textualism to work out more details and address a greater number of contingencies in case a conflict arises might be quite minimal.

³⁷⁴ Cohen, *supra* note 272, at 134; accord Posner, *supra* note 195, at 1600.

latter approach best minimizes overall expenses. But the actual answer to the question is unknown.³⁷⁵

3. *Enforcement Costs*

Perhaps the signature argument textualists offer in favor of their interpretive approach is that enforcement costs are higher under contextualism than under textualism.³⁷⁶ Both courts and scholars have regularly pressed this claim.³⁷⁷ The claim has two components: enforcement costs are greater in a contextualist regime because (1) there are more lawsuits, and (2) the lawsuits that are filed last longer.³⁷⁸

Start by recalling that a much broader range of material is relevant in deciding whether a contract is ambiguous under contextualism. Textualism recognizes only patent ambiguities. Thus, when attempting to convince the court that a contract is reasonably susceptible to more than one meaning, a party may rely solely upon the language within the four corners of the agreement. The judge is barred from considering any other evidence.³⁷⁹ Contextualism recognizes both patent and latent ambiguities. Accordingly, a party may use the language of the agreement *as well as* preliminary negotiations, courses of performance and dealing, trade usages, and other aspects of the surrounding context in attempting to establish the existence of an ambiguity.³⁸⁰ Next, recall that textualists maintain that extrinsic

³⁷⁵ Note that even if we could be reasonably certain about which approach best reduces overall costs, we would also need to know the magnitude of the difference in order to weigh cost reduction against interpretive accuracy and any other factors under consideration.

³⁷⁶ See MITCHELL, *supra* note 316, at 108 (“The first and most obvious reason for confining a court’s enquiry to the four corners of the agreement relates to the possible costs involved in the contextual approach.”).

³⁷⁷ See *id.* (noting that the costs of the contextualists approach, including enforcement costs, “has been a particular concern of some judges”); Cohen, *supra* note 272, at 133 (“Law and economics scholars often argue that contextualism is associated with higher litigation costs than textualism.”).

³⁷⁸ As used here, “last longer” denotes more than the mere passage of time. Rather, it means moving into later stages of the litigation process, with each stage requiring new activities that entail the expenditure of resources. Note also that the number of lawsuits filed and the length of those suits are only indirect measures of enforcement costs. Directly quantifying such costs would require analyzing party and court expenditures on items like attorney’s fees, taxable costs, filing fees, and time spent by the judiciary addressing interpretation disputes. Nonetheless, there appears to be almost universal agreement that the number of actions brought and how long those actions last are sufficient proxies.

³⁷⁹ See *supra* notes 286–88, 300–01 and accompanying text.

³⁸⁰ See *supra* notes 295–97, 302–03, and accompanying text.

evidence has many problematic features: it is frequently unreliable, contradictory, and/or ambiguous.³⁸¹ This means that contextualism both dramatically increases the *quantity* of relevant interpretive material that courts must consider at the ambiguity stage *and* reduces the *quality* of the material that goes into the ambiguity determination. These two features of contextualism raise enforcement costs from the textualist baseline through four pathways—in addition to the incentives that contextualism create to draft worse contracts discussed in the prior section.³⁸²

First, it is much more difficult for contextualist courts to decide contract interpretation cases on the pleadings. Since a party is entitled to argue that an agreement is ambiguous via extrinsic evidence, the court generally must permit discovery so that such evidence can be gathered. Therefore, the ambiguity determination typically can be made no earlier than at the summary judgment stage.³⁸³ Second, it is easier to establish that a contract is ambiguous when extrinsic evidence is available because the parties have more material out of which to craft reasonable constructions of the operative language.³⁸⁴ And contextualism incentivizes parties to invest

³⁸¹ See *supra* notes 333–34 and accompanying text.

³⁸² See *supra* notes 372–73 and accompanying text.

³⁸³ See *Bank v. Truck Ins. Exch.*, 51 F.3d 736, 737–38 (7th Cir. 1995) (Posner, J.) (observing that contextualist interpretation “makes it difficult to decide contract cases on the pleadings”); see also MITCHELL, *supra* note 316, at 110 (observing that contextualist litigation is expensive because “the relevant ‘context’ has to be established”); *id.* at 63 (explaining that judges are justifiably concerned that contextualism adds to the costs and delays of litigation because they must consider the context before deciding what the contract means). Indeed, some courts have held that an interpretation case may *never* be decided on the pleadings under contextualism. See, e.g., *A. Kemp Fisheries, Inc. v. Castle & Cooke, Inc.*, 852 F.2d 493, 497 n.2 (9th Cir. 1988) (stating that under California law, “courts may not dismiss on the pleadings when one party claims that extrinsic evidence renders the contract ambiguous”). But unless a jurisdiction has adopted the “private code” theory endorsed by the *Restatement (Second)*, see *supra* note 314, this goes too far. Remember that the language of the agreement must be reasonably susceptible to the proffered interpretation. Therefore, a contextualist court may bar discovery and decide the case on the pleadings when the construction advanced by one party is so inconsistent with the text that no set of extrinsic evidence could make the text reasonably susceptible to the asserted reading. See *Hervey v. Mercury Cas. Co.*, 110 Cal. Rptr. 3d 890, 895–96 (Cal. Dist. Ct. App. 2010) (“Although parol evidence may be admissible to determine whether the terms of a contract are ambiguous, it is not admissible if it *contradicts* a clear and explicit provision.”) (emphasis added) (further explaining the circumstances in which contract interpretation cases can be decided on the pleadings).

³⁸⁴ See Goldstein, *supra* note 281, at 75 (“The plain meaning rule allows more sophisticated parties to hide behind carefully worded contracts of adhesion without fear that the circumstances surrounding the contract might intrude.”).

heavily in the search for evidence that can support their preferred construction of the contract.³⁸⁵

Given these two points, lawsuits will generally last longer when courts employ contextualist methodology; more cases will reach discovery, summary judgment, and trial.³⁸⁶ In addition, the parties are more likely to file a lawsuit to begin with since those challenging the apparently clear terms of a contract stand a better chance of surviving the ambiguity stage and making it to a jury than if the courts use textualism.³⁸⁷

Third, the quantity and quality of the extrinsic evidence available under contextualism makes it more difficult to predict how judges will

³⁸⁵ See MITCHELL, *supra* note 316, at 113 (“A further problem is that much reliance on context may be done strategically—the problem of ‘threshing through the undergrowth’ for the chance remark upon which to build a case. The suspicion is often raised of the strategic reliance on context to sanction an escape from a bad bargain[,] . . . One may use the ‘context’ to seek an unbargained for advantage in imposing terms after the parties are in a contractual relationship, even in circumstances where the written terms appear relatively complete.”); Cohen, *supra* note 272, at 133 (“For example, allowing more contextual evidence may encourage parties to spend more on litigation because the marginal benefit of expenditures to develop such evidence is higher than under a textualist regime.”); Katz, *supra* note 357, at 530 (“Under a regime of substantive interpretation, for instance, parties may be tempted to invest substantial resources in litigation in order to maximize the chance of a favorable outcome.”); *id.* at 531 (“Formality, by limiting the scope for *ex post* interpretive disputes, probably reduces the marginal productivity of litigation expenditure, and thus reduces the amount of such expenditure.”). See also Steuart v. McChesney, 444 A.2d 659, 663 (Pa. 1982) (“Likewise, resort to the plain meaning of language hinders parties dissatisfied with their agreement from creating a myth as to the true meaning of the agreement through subsequently exposed extrinsic evidence.”).

³⁸⁶ See Ward v. Intermountain Farmers’ Ass’n, 907 P.2d 264, 269–70 (Utah 1995) (Russon, J., concurring) (“When a motion for summary judgment can be defeated merely by the opposing party’s affidavit averring that an otherwise clear contract provision was intended to mean something different, attorneys will discontinue the futility of composing summary judgment motions, and every contract dispute will be formally resolved only through trial.”); Schwartz & Scott, *supra* note 9, at 963 (“The plain meaning rule operates in tandem with a hard parol evidence rule to reduce expected adjudication costs. If the contract is fully integrated, and if contractual terms are facially clear, then the dispute can be resolved at summary judgment.”); Spigelman, *supra* note 328, at 413 (“Furthermore, the length and cost of the process is increased [by contextualism].”). See also BURTON, *supra* note 9, § 4.6.2, at 147 (“For both parties and others, investigating the parties’ subjective intentions can be costly, if such investigations are possible without rights to discovery and perhaps even then.”); Shavell, *supra* note 322, at 311 (observing that “extrinsic evidence . . . is very costly to consider (especially because of the tendency of the parties to contest negotiating history, oral statements, course of dealing).”).

³⁸⁷ See FDIC v. W.R. Grace & Co., 877 F.2d 614, 621 (7th Cir. 1989) (Posner, J.) (“The older view, sometimes called the ‘four corners’ rule, . . . tends to cut down on the amount of litigation.”). Note that most textualists would probably contend that contextualist interpretation also increases the number of disputes over contractual meaning that do *not* result in a lawsuit. Such disputes are another type of enforcement cost. But this argument does not appear to play a significant role in the secondary literature or caselaw.

resolve the ambiguity question.³⁸⁸ Parties do not know which contractual language or other evidence the court is likely to find dispositive.³⁸⁹ And, if no lawsuit has been filed yet, neither party will even have access to all of the materials the judge is going to consider since discovery will not have started. Fourth, as noted above, interpretation cases are more likely to reach trial under contextualism than under textualism.³⁹⁰ And jury trials are considered notoriously difficult to predict.³⁹¹ It is generally accepted that

³⁸⁸ See GOLDBERG, *supra* note 336, at 162 (“The danger of a *Nanakuli-Columbia Nitrogen* [contextualist] interpretative strategy is that parties will be frustrated in trying to devise the terms of their agreement, and they will have little confidence in their ability to predict the outcomes if their disputes do end in litigation.”); FARNSWORTH, *supra* note 270, § 7.12, at 465 (“The restrictive view is defended on the grounds that it . . . gives predictability in the interpretation of commonly used terms.”); KNIFFIN, 5 CORBIN ON CONTRACTS, *supra* note 281, § 24.7, at 53 (noting that various “judges have expressed the view that discarding the plain meaning rule would interfere with predictability and uniformity in interpretation of contracts . . .”); MCMEEL, *supra* note 305, § 1.107 (observing that after English courts adopted contextualism, “fears were expressed” that this would “generate greater uncertainty in the context of commercial transactions”); MITCHELL, *supra* note 316, at 92 (“The weakness of contextualism is its unpredictability.”); Posner, *supra* note 312, at 562 & n.14 (“Courts that support hard-PER argue that this rule increases commercial certainty by enabling parties to predict the promises that courts will enforce.”) (collecting authorities); Goldstein, *supra* note 281, at 76 (“Most problematically, by looking to evidence of the parties’ subjective intent, rather than the shared and public meaning of terms, the context rule undermines the usefulness of contracts as tools to predictably constrain another party’s behavior.”); Kniffin, *supra* note 304, at 100 n.30 (“The plain meaning rule is intended to avoid unnecessary expenditures of judicial resources and to further predictability; the court avoids examining extrinsic evidence when the court is certain of the meaning of a disputed term, . . .”); Spigelman, *supra* note 328, at 412 (arguing that “the general use of extrinsic materials” undermines certainty “in contracts between commercial parties” and results “in an increase in the cost of commercial dispute resolution”); see also *Herson v. Gibraltar Building & Loan Ass’n*, 864 F.2d 848, 853 (D.C. Cir. 1989) (“Nonetheless, it is fundamentally important that parties be able to rely on the explicit language of written contracts. The public interest in certainty and finality is too critical to allow every agreement to be subjected to collateral attack.”); *W.W.W. Assoc., Inc. v. Giancontieri*, 566 N.E.2d 639, 643 (N.Y. 1990) (“An analysis that begins with consideration of extrinsic evidence of what the parties meant, instead of looking first to what they said and reaching extrinsic evidence only when required to do so because of some identified ambiguity, unnecessarily denigrates the contract and unsettles the law.”).

³⁸⁹ See MITCHELL, *supra* note 316, at 91 (“Parties, and their lawyers, may . . . have little idea of what outcomes [contextualism] may lead to, since they may be unaware of what particular context, and contextual material, is regarded as controlling.”).

³⁹⁰ See *supra* note 386 and accompanying text.

³⁹¹ Valerie P. Hans & Theodore Eisenberg, *The Predictability of Juries*, 60 DEPAUL L. REV. 375, 375 (2011) (“The jury is said to be the least predictable of the decision makers in the legal system.”). See, e.g., Dru Stevenson, *The Function of Uncertainty Within Jury Systems*, 19 GEO. MASON L. REV. 513, 513 (2012) (“Indeed, current jury selection methods all but guarantee that jury trial outcomes are uncertain and unpredictable.”); Byron G. Stier, *Another Jackpot (In)justice: Verdict Variability and Issue Preclusion in Mass Torts*, 36 PEPP. L. REV. 715, 720 (2009) (“Evidence of variability conforms to lawyers’ long-held beliefs about the unpredictability of trial. Juries may well deliver verdicts that

adjudicative uncertainty increases litigation.³⁹² Therefore, the uncertainty contextualism creates at the ambiguity stage and through the greater number of trials increases the likelihood that parties will file a lawsuit.³⁹³ In addition, because uncertainty reduces the probability of settlement, contextualism tends to lengthen any interpretation litigation that is commenced.³⁹⁴

When courts advance the claim that contextualism increases enforcement costs, they often do so with considerable stridency and with

substantially differ, though based on identical facts.”); *but see* Hans & Eisenberg, *supra*, at 379–80 (reviewing the literature and concluding that juries are “generally predictable in the sense that we know what particular factors will lead to plaintiff verdicts and substantial compensatory and punitive damages awards”).

³⁹² James M. Fischer, *Discretion and Politics: Ruminations on the Recent Presidential Election and the Role of Discretion in the Florida Presidential Election Recount*, 69 U. CIN. L. REV. 807, 836 n.99 (2001) (“The view that uncertainty increases litigation costs appears to be generally held.”); Siegelman & Donohue, *supra* note 153, at 1148 (“A substantial literature also indicates that uncertainty about the likely outcome of a trial will diminish the chance that the case will be settled.”); *see, e.g.*, Priest & Klein, *supra* note 154, at 45 (“Substantial uncertainty over the outcome of individual trials, of course, will lead in general to high rates of litigation . . .”).

³⁹³ MITCHELL, *supra* note 316, at 112 (“A related problem is that litigation over terms and obligations is actually *encouraged* (and hence costs incurred) by courts adopting a contextual approach, . . .”) (emphasis in original); LINZER, 6 CORBIN ON CONTRACTS, *supra* note 310, § 25.15[A], at 190 (nothing that a “common argument against the loosening of the parol evidence rule” is that “the use of extrinsic evidence to show ambiguity could open floodgates” to more litigation); *see also* BURTON, *supra* note 9, § 1.1.2, at 7 (explaining that “predictability encourages performance [and] discourages disputes . . .”).

³⁹⁴ *See* Cohen, *supra* note 272, at 133 (“Alternatively, allowing contextual evidence may undermine certainty and therefore make settlement less likely.”); *see also* BURTON, *supra* note 9, § 1.1.2, at 7 (explaining that “predictability . . . fosters settlement”); GOLDBERG, *supra* note 336, at 163 (“The role of the formal law, in this view, is to provide an anchor. If the litigation outcome is relatively certain, it provides a clear base point for negotiating a settlement.”); Katz, *supra* note 357, at 531 (“To the extent that it [textualism/formalism] conditions the outcome of litigation on publicly available information, and reduces the variations of litigant’s expectations regarding that outcome, it probably also encourages settlement.”).

Note that textualists abroad sometimes argue that discovery and trials last longer under contextualism and that trials are harder to predict. *See, e.g.*, MacLauchlan, *supra* note 10, at 36 (explaining that the English case adopting contextualism “was seen as a recipe for a further increase in the already substantial cost of the discovery process and the lengths of trials”); MCMEELE, *supra* note 305, §§ 1.107, 1.109, 1.110 (same). That might be true under the versions of textualism used in other countries. But as I explained previously, textualist and contextualist courts in the United States generally concur on the evidence that may be used at trial after the court has determined that an ambiguity exists. *See supra* note 298 and accompanying text. Accordingly, there generally should be no difference in the length of discovery, the length of trials, or the predictability of trials under the two approaches. Some scholars, however, have argued that a more limited range of evidence should be used at trial. *See, e.g.*, BURTON, *supra* note 9, ch. 6, at 193; *id.* § 6.1.3.

language that tends to conflate the various arguments just discussed. An excellent example can be found in *Trident Center v. Connecticut General Life Ins. Co.*³⁹⁵ There, Judge Kozinski set forth his now famous assault on California contract interpretation doctrine.³⁹⁶ Taking aim at *Pacific Gas & Electric Co. v. G.W. Thomas Drayage & Rigging Co.*,³⁹⁷ the watershed California Supreme Court decision that paved the way for modern acceptance of the contextualist approach, he wrote the following:

Pacific Gas casts a long shadow of uncertainty over all transactions negotiated and executed under the law of California. As this case illustrates, even when the transaction is very sizeable, even if it involves only sophisticated parties, even if it was negotiated with the aid of counsel, even if it results in contract language that is devoid of ambiguity, costly and protracted litigation cannot be avoided if one party has a strong enough motive for challenging the contract. While this rule creates much business for lawyers and an occasional windfall to some clients, it leads only to frustration and delay for most litigants and clogs already overburdened courts.³⁹⁸

Comparable statements abound in the caselaw.³⁹⁹

³⁹⁵ *Trident Center v. Connecticut General Life Ins. Co.*, 847 F.2d 564 (9th Cir. 1988).

³⁹⁶ See Linzer, *supra* note 311, at 814 (referring to Judge Kozinski's Trident opinion as "famous").

³⁹⁷ 442 P.2d 641 (Cal. 1968).

³⁹⁸ *Trident Center*, 847 F.2d at 569.

³⁹⁹ Indeed, California Supreme Court Justice Stanley Mosk made essentially the same arguments in one of the cases commonly associated with *Pacific Gas* that was decided later that same year:

Given two experienced businessmen dealing at arm's length, both represented by competent counsel, it has become virtually impossible under recently evolving rules of evidence to draft a written contract that will produce predictable results in court. The written word, hereto for deemed immutable, is now at all times subject to alteration by self-serving recitals based upon fading memories of antecedent events. This, I submit, is a serious impediment to the certainty required in commercial transactions.

Delta Dynamics, Inc. v. Arioto, 446 P.2d 785, 789–90 (Cal. 1968) (Mosk, J., dissenting). Another excellent example can be found in *Steuart v. McChesney*, where the Pennsylvania Supreme Court wrote this:

Accordingly, the plain meaning approach enhances the extent to which contracts may be relied upon by contributing to the security of belief that the final expression of *consensus ad idem* will not later be construed to import a meaning other than that clearly expressed. . . . Likewise, resort to the plain meaning of language hinders parties dissatisfied with their agreement from creating a myth as to the true meaning of the agreement through subsequently exposed extrinsic evidence. Absent the plain meaning rule, nary an agreement could be conceived, which, in the event of a party's later disappointment with his stated bargain, would not be at risk to having its true meaning obfuscated under the guise of examining extrinsic evidence of intent. Even if the dissatisfied party in good faith believed that the agreement, as manifest, did not express the

Not surprisingly, some advocates of contextualism counter that textualism is the approach with higher enforcement costs. They offer several arguments in defense of this position. First, because textualism prohibits the review of extrinsic evidence when determining whether an agreement is ambiguous, the principal inputs at that stage are (1) the contract, and (2) the judge.⁴⁰⁰ But judges “come from a variety of backgrounds—private law practice, government service, business, academia—and their fields of experience represent an even wider variance.”⁴⁰¹ Such differences can lead judges to reach disparate conclusions regarding the same contractual language. Indeed, “[a]ppellate courts’ reviews of four corner determinations are often arbitrary and extremely subjective.”⁴⁰² Critically, the parties will not know which trial judge is going to interpret their contract until a lawsuit is filed. Nor will they know which appellate judges are going to be assigned to the case if the dispute reaches a higher court.⁴⁰³ This makes it immensely difficult for parties to predict the results of ambiguity decisions in textualist jurisdictions.⁴⁰⁴ Such uncertainty increases the number of lawsuits and hinders settlements.⁴⁰⁵

consensus ad idem, his post hoc judgment would be inclined to be colored by belief as to what should have been, rather than what strictly was, intended.

444 A.2d 659, 663 (Pa. 1982) (citations omitted).

⁴⁰⁰ See *supra* notes 285–88 and accompanying text. Other textualist inputs include dictionaries, the rules of grammar, the canons of construction, and any arguments made by the parties. See *supra* note 286.

⁴⁰¹ *Mellon Bank, N.A. v. Aetna Business Credit*, 619 F.2d 1001, 1011–12 n.12 (3d Cir. 1980).

⁴⁰² LINZER, 6 CORBIN ON CONTRACTS, *supra* note 310, § 25.14[B], at 163.

⁴⁰³ See Burton, *supra* note 10, at 357.

⁴⁰⁴ See LINZER, 6 CORBIN ON CONTRACTS, *supra* note 310, § 25.14[B], at 163 (explaining that “it is difficult to predict how the appellate courts will read words claimed to be ambiguous” in textualist states). See also *Mellon Bank*, 619 F.2d at 1010 (“If each judge simply applied his own linguistic background and experience to the words of a contract, contracting parties would live in a most uncertain environment.”); Burton, *supra* note 10, at 353 (“So, because the parties will not know who the judge will be if litigation ensues, OCI [Objective Contextual Interpretation] better contains pre-litigation costs. OCI better enables the parties to forecast an adjudicatory result when they draft a contract, consider whether to perform or breach, decide whether to challenge the other party’s performance, attempt to settle a dispute, and plan for litigation.”).

⁴⁰⁵ Burton, *supra* note 10, at 357 (“Due to the uncertainties [regarding which judge the parties will appear before], moreover, both trial and appellate proceedings would proliferate, the latter because appellate judges will have different backgrounds from both the trial judges and from one another.”); *id.* at 353 (quoted in footnote 404, *supra*).

Second, in many cases, extrinsic evidence can show that a contract that appears ambiguous on its face is actually perfectly clear. “[M]ost words have several meanings in the abstract (acontextually). With a little context, we may know easily which meaning is apt.”⁴⁰⁶ As a result, contextualism may well decrease the number of lawsuits in which the court finds the contract to be ambiguous, reducing the length of these actions and promoting greater certainty.⁴⁰⁷

Third, under textualism, parties have an incentive to write longer and more complete contracts.⁴⁰⁸ Since the court will not consider anything but the text of the agreement during the first stage of the interpretive process, any understanding of the parties not expressly reduced to writing will be inoperative when their contract is facially unambiguous. The parties cannot rely upon extrinsic evidence, such as usages of trade, to explain, supplement, or qualify the contractual terms. But longer written contracts are more complex. And “[g]reater complexity can in fact lead to more litigation, as the chance that terms will conflict or support alternative conduct increases,”⁴⁰⁹ making a finding of ambiguity more likely.

I will add a fourth contextualist argument of my own. In my experience, the average person, whether a consumer or in business, is angered or even outraged when a counter-party insists on the strict application of unambiguous contractual language that appears to conflict with the prior contextual understanding of the parties. This is especially true when the counter-party stated during preliminary negotiations that the relevant language was of no consequence or would not be relied upon should conditions change or a dispute arise. Such conduct may infuriate a

⁴⁰⁶ See BURTON, *supra* note 9, § 6.1.2.2, at 210.

⁴⁰⁷ See Posner, *supra* note 312, at 562 (“Courts that support soft-PER argue that soft-PER increases commercial certainty by allowing judges and juries to consider all relevant evidence.”); LINZER, 6 CORBIN ON CONTRACTS, *supra* note 310, § 25.14[B], at 163 (“Given the arbitrariness of the decisions [in Texas], the courts would have been better off making them with the additional information offered by the rejected extrinsic evidence.”); MacLauchlan, *supra* note 10, at 35 (explaining that allowing the admission of prior negotiations could reduce interpretive uncertainty since it will sometimes reveal “that the parties formed a common intention as to the meaning of the words in dispute”).

⁴⁰⁸ See Cohen, *supra* note 272, at 134.

⁴⁰⁹ *Id.* For an excellent example, see *Doctor’s Associates, Inc. v. Dupree*, 745 N.E.2d 1270, 1281 (Ill. App. Ct. 2001) (settlement agreement contained both a general release extinguishing claims against numerous third-party beneficiaries and a clause stating that the contract was not intended to provide contractual rights to any third-party beneficiaries).

consumer or business sufficiently to motivate them to sue, or to resist to the point that the other side is compelled to file an action. If textualism incentivizes parties to stand on language that is inconsistent with the other side's reasonable expectations more than contextualism⁴¹⁰—if textualism promotes behavior that increases the likelihood that contractual partners will become frustrated and accept going to court—then this may be another pathway through which textualism increases litigation.⁴¹¹

4. Other Policy Arguments

Scholars have offered a number of policy arguments regarding contract interpretation that do not fit squarely into the accuracy/transaction costs/enforcement costs framework set forth above. This subpart briefly discusses the most important of those arguments.

Textualism is frequently defended on the ground that businesses prefer that method of construction.⁴¹² This view finds support in the work of Lisa

⁴¹⁰ Mitchell explains that textualism is subject to this type of abuse: “[A] party may strategically seek an advantage by relying on the strict words of a contract while knowing that the documents did not reflect the parties’ joint understanding . . .” MITCHELL, *supra* note 316, at 113.

⁴¹¹ One final note is in order. As Parts IV.B.1. and IV.B.3 demonstrate, both sides in the debate tend to claim superiority on accuracy and enforcement costs. However, many commentators conceptualize the choice between the two interpretive approaches as one involving trade-offs. These scholars are willing to concede that the other side has the better argument on at least one of the three, key issues; but they contend that their own side is still superior because it is much stronger on the remaining dimension(s). For example, as explained in Part IV.B.2., textualists generally acknowledge that contextualism lowers transaction costs; but they believe that it raises enforcement costs by a higher amount. *See also* Cohen, *supra* note 272, at 133 (observing that a “number of scholars have argued that the optimal contract rules of interpretation and implied terms are determined by the tradeoff between ex ante negotiation and drafting costs and ex post litigation costs”); *supra* note 328 (setting forth several secondary sources that use the trade-off framework). Similarly, some textualists concede that contextualist interpretation is superior on the issue of accuracy, but that such accuracy is not worth the increased enforcement costs. *See, e.g.*, Schwartz & Scott, *supra* note 9, at 933 (“Moreover, we concede that a court is more likely to make an accurate interpretation if it sees more evidence, but we argue that sometimes accuracy is not worth the costs of achieving it.”); *compare* LINZER, *supra* note 310, § 25.4, at 39 (conceding that textualism lowers enforcement costs, but arguing that contextualism’s greater accuracy is worth the price).

⁴¹² STEPHEN A. SMITH, CONTRACT THEORY 276 (2004) (explaining that textualism is grounded on the assumption that those drafting the contract intended that the terms “be read narrowly and literally”); “[t]he context of commercial drafting, in other words, is one that asks the reader to ignore the context outside of the physical document”); *see, e.g.*, Spigelman, *supra* note 328, at 429 (“Nevertheless, the idea that an arbitrator or a judge would be called upon to determine the true intention of the parties by going beyond the written contract to encompass anything which disputing parties can relevantly

Bernstein. Bernstein surveyed the contract interpretation practices of merchant courts in the private legal systems of the grain and feed industry and the cotton industry. She found that the tribunals in both industries use interpretive methods that are substantially textualist in nature.⁴¹³ This supports the conclusion that businesses favor textualism.⁴¹⁴ A common explanation for this preference is that merchants prefer the certainty of transaction costs to the uncertainty of enforcement costs.⁴¹⁵ Alternatively, Bernstein hypothesizes that businesses “do not necessarily want the relationship-preserving norms they follow in performing contracts to be used by third-party neutrals to decide cases when they are in an end-game

imagine, would be regarded by most parties, at the time of formation of the contract, to constitute a commercial disaster.”); *id.* at 412 (“In this paper, I will be concerned with contracts between commercial parties—not with consumers.”); *see also* Schwartz & Scott, *supra* note 9, at 932 (“[B]oth the available evidence and prevailing judicial practice support the claim that sophisticated parties prefer textualist interpretation.”).

⁴¹³ *See* Lisa Bernstein, *Merchant Law in a Merchant Court: Rethinking the Code’s Search for Immanent Business Norms*, 144 U. PA. L. REV. 1765, 1769–70 (1996) [hereinafter Bernstein, *Merchant Law*] (finding that National Grain and Feed Association arbitrators, “despite their industry expertise, take a formalist approach to adjudication”; these arbitrators are “reluctant to look to” course of performance, course of dealing, and usage of trade in construing agreements and “do not permit these considerations to vary either trade rules or written contract provisions”); Lisa Bernstein, *Private Commercial Law in the Cotton Industry: Creating Cooperation through Rules, Norms and Institutions*, 99 MICH. L. REV. 1724, 1735 (2001) [hereinafter Bernstein, *Private Commercial Law*] (“[D]espite the fact that cotton arbitrators are chosen for their industry expertise, they use a relatively formalistic adjudicative approach that gives little explicit weight to elements of the contracting context.”); *id.* at 1735–36 (noting (1) that cotton trade rules do not make course of performance, course of dealing, and usage of trade relevant to the interpretation of contracts, (2) that cotton arbitrators “are reluctant to take [course of performance and course of dealing] into account,” and (3) that “references to custom or usage in [cotton arbitration] opinions are extraordinarily rare”); *see also* Lisa Bernstein, *The Questionable Empirical Basis of Article 2’s Incorporation Strategy*, 66 U. CHI. L. REV. 710, 713–17, 751–53 (1999) [hereinafter Bernstein, *Questionable Empirical Basis*] (concluding, based on an empirical study of the hay, grain and feed, textiles, and silk industries, that usages of trade rarely exist in the form contemplated by the Uniform Commercial Code).

⁴¹⁴ *See* MITCHELL, *supra* note 316, at 103 (explaining that Bernstein’s studies “suggest that some contractors, in some circumstances, prefer a more formalist approach to be taken”); Geoffrey P. Miller, *Bargains Bicoastal: New Light on Contract Theory*, 31 CARDOZO L. REV. 1475, 1477 (2010) (“Bernstein’s work suggests that industry actors, when given the freedom to devise their own procedures, opt for a system of rules much like that predicted in Schwartz and Scott’s [formalist] theory.”).

⁴¹⁵ *See, e.g.*, Cohen, *supra* note 272, at 134 (“Judge Posner posits that the four corners rule is based on the assumption that parties prefer ex ante contracting to the expense and uncertainty of a jury trial.”) (citing Posner, *supra* note 195, at 1602–03); MITCHELL, *supra* note 316, at 91 (“Given this uncertainty [caused by contextualism], the possibility presents itself that some parties may prefer a more formal interpretative method[.]”).

situation.”⁴¹⁶ They favor a system that permits them to deploy contextualist methodologies when negotiating informally with a counterparty “while at the same time retaining their right to insist on strict adherence to the terms of their written contract if their relationship breaks down.”⁴¹⁷ If textualism is in fact preferred by merchants, then there is good reason to believe that it maximizes the value of commercial agreements since businesses are in the best position to determine the optimal trade-off of accuracy, transaction costs, and enforcement costs in their dealings.⁴¹⁸

Contextualism is also defended on the ground that contracting parties favor that approach. Indeed, the drafters of the Uniform Commercial Code justified the statute’s highly contextualist interpretive rules, in part, on the belief that merchants intend and understand trade usage and other aspects of the commercial context to be essential components of business agreements.⁴¹⁹ And subsequent commentators, particularly those from the

⁴¹⁶ Bernstein, *Merchant Law*, *supra* note 413, at 1770; *see also id.* at 1770–71, 1796–1820 (elaborating on this argument).

⁴¹⁷ Bernstein, *Private Commercial Law*, *supra* note 413, at 1780–81. *See also* Theodore Eisenberg & Geoffrey P. Miller, *The Flight to New York: An Empirical Study of Choice of Law and Choice of Forum Clauses in Publicly-Held Companies’ Contracts*, 30 CARDOZO L. REV. 1475, 1475–77, 1490, 1504 (2009) (empirical study of choice-of-law and choice-of-forum provisions in 2,882 contracts reflecting major transactions and attached as exhibits to SEC filings; far more contracts opted for New York law or a New York forum than for California law or a California forum); Miller, *supra* note 414, at 1477–79 (2010) (concluding, based on the study described in the prior parenthetical, that “the verdict of thousands of sophisticated parties whose incentives are to maximize the value of contract terms . . . is that New York’s formalist rules win out over California’s contextualist approach”). Eisenberg and Miller’s work has been cited for the proposition that businesses favor textualist interpretation. *See* Schwartz & Scott, *supra* note 9, at 956–57. But Miller’s article identified roughly seventeen doctrinal differences between New York and California, only one of which was contract interpretation. *See* Miller, *supra*, at 1481–1522. Thus, it is far from clear that differences in interpretive regimes played an important role in the choice of law and forum decision-making that Eisenberg and Miller studied. *See also* Burton, *supra* note 10, at 348–49 n.64 (essentially identifying the same problem with relying upon Eisenberg and Miller’s study to defend the claim that firms prefer textualism).

⁴¹⁸ *See* Schwartz & Scott, *supra* note 9, at 930 (explaining, in a closely related context, that “parties are better informed than courts about benefits and costs, so parties commonly have a comparative advantage over courts in making the required tradeoffs”).

⁴¹⁹ *See* LARRY A. DIMATTEO ET AL., VISIONS OF CONTRACT THEORY: RATIONALITY, BARGAINING, AND INTERPRETATION 160 (2007); Bernstein, *Questionable Empirical Basis*, *supra* note 413, at 746–47. Note that Bernstein’s empirical studies are directed at challenging the approach of the Uniform Commercial Code, *see, e.g.*, Bernstein, *Merchant Law*, *supra* note 413, at 1766, including the Code’s assumption about the preferences of merchants, *see, e.g.*, Bernstein, *Questionable Empirical Basis*, *supra*, at 751–52.

relational contracting school of thought, have endorsed a similar view.⁴²⁰ One possible explanation for such a preference is that parties would rather reduce transaction costs than enforcement costs. That is because transaction costs are incurred with certainty and in the present while enforcement costs are incurred rarely and well into the future.⁴²¹ Note also that scholars have challenged the claim that Bernstein's work supports the conclusion that businesses prefer textualism. For example, her findings might be explained by the fact that the arbitrators serving on commercial tribunals have extensive industry experience and thus understand the relevant customs and usages. As a result, submitting additional contextual evidence in cases before these courts will not sufficiently reduce the risk of an interpretive error to justify the increased enforcement costs. But generalist judges and juries lack the commercial background of industry experts. Accordingly, the fact that firms prefer textualist interpretation when appearing before private commercial tribunals does not entail that they would have the same preference in a lawsuit heard in state or federal court where the accuracy-enhancing effects of extrinsic evidence might be worth the added cost.⁴²²

A third possibility is that the interpretive preferences of contracting parties are heterogeneous.⁴²³ For example, some commentators have hypothesized that risk-averse parties favor contextualism while risk-neutral parties prefer textualism.⁴²⁴

The possibility of diverse preferences raises another issue. Much of the interpretation debate presumes that one school of thought is inherently

⁴²⁰ See, e.g., Robert A. Hillman, *Regulating Contracts by Hugh Collins*, 27 J.L. & Soc'y 338, 343 (2000) (summarizing the relational contracting theory of Ian Macneil).

⁴²¹ MITCHELL, *supra* note 316, at 109–10 (citing Schwartz & Scott, *supra* note 338, at 585).

⁴²² Katz, *supra* note 357, at 526–27; see also Burton, *supra* note 10, at 348 n.64 (contending that Bernstein's studies "cannot be easily generalized" beyond the specific industries addressed in her work). Bernstein recognizes the possibility raised by Katz. See Bernstein, *Private Commercial Law*, *supra* note 413, at 1735 n.57 ("Given the expertise of these arbitrators, however, these considerations [namely, the commercial context] may enter the moving papers and/or influence the arbitrators' decision-making processes in ways too subtle to detect."); see also Bernstein, *Questionable Empirical Basis*, *supra* note 413, at 716 n.18 ("The opinions produced by merchant tribunals reveal that arbitrators' background knowledge of the trade may enable them to better assess the credibility of testimony and may give them a better understanding of the types of evidence that ought to be submitted.").

⁴²³ See Schwartz & Scott, *supra* note 9, at 930 ("[P]arty preferences over interpretive rules are heterogeneous.").

⁴²⁴ See Cohen, *supra* note 272, at 143 (discussing such theories).

superior.⁴²⁵ But perhaps the optimal interpretive regime varies based on the circumstances, such as the type of contract and the identity of the parties. This view is increasingly popular in the literature.⁴²⁶ For example, some courts and scholars believe that textualism should be used to construe agreements between businesses, while contextualism is the best approach for contracts involving consumers.⁴²⁷ Others have developed sophisticated models to explain when each interpretive approach is likely to be superior.⁴²⁸

The last argument worth discussing here is the contention that contracting parties should be entitled to choose which interpretive rules govern their agreement. In other words, contract interpretation doctrine should consist of default rules, like the bulk of the rest of contract law, rather than mandatory rules.⁴²⁹ This position has strong support among academic commentators. Indeed, it is apparently endorsed by most economists who have studied the subject.⁴³⁰ Perhaps the leading advocates of this view in the legal academy are Alan Schwartz and Robert Scott. Their

⁴²⁵ Adam B. Badawi, *Interpretive Preferences and the Limits of the New Formalism*, 6 BERKELEY BUS. L.J. 1, 4 (2009).

⁴²⁶ See, e.g., *id.* at 5 (“This article argues that the desirability of an interpretive regime depends, at least to some degree, on the attributes of the underlying transactions and not solely on the independent merits of formal or contextual interpretation.”); Katz, *supra* note 357, at 538 (explaining that this essay presents a “basic framework” for determining “in which contexts and for which parties formalism is most useful and in which contexts and for which parties a substantive approach is most useful”); Kostritsky, *supra* note 317, at 44 (“This article argues that it is wrong to think that courts must make a dichotomous choice always to prefer extrinsic evidence or always to exclude it. Sometimes the appropriate interpretive methodology should explicitly forego extrinsic evidence while at other times it should embrace extrinsic evidence.”).

⁴²⁷ Katz, *supra* note 357, at 538; see also Bernstein, *Merchant Law*, *supra* note 413, at 1820–21 & n.168 (acknowledging that the case for contextualist interpretation is “far stronger in merchant-to-consumer transactions” than in merchant-to-merchant transactions).

⁴²⁸ See *supra* note 426; see also Hermalin et al., *supra* note 9, at 90–91 (setting forth “some general rules of thumb” that identify the circumstances in which textualist interpretation is more efficient than contextualist interpretation, and vice versa) (“It follows from these heuristic principles that substantive interpretation is relatively more valuable to small and infrequent traders . . . [while] large and experienced traders should prefer their contracts to be governed by relatively formalistic rules of interpretation.”).

⁴²⁹ “Default rules are rules that parties can contract around, whereas mandatory rules apply regardless of the parties’ intentions.” Cohen, *supra* note 272, at 135.

⁴³⁰ See *id.* (also summarizing the general case for default rather than mandatory terms). *But cf.* Schwartz & Scott, *supra* note 9, at 939 (“Just about everyone who creates, applies, or analyzes the interpretive rules believes that they should be mandatory.”) (probably referring to people working in the legal field rather than economists).

argument that contracts between businesses⁴³¹ should be governed by default interpretation rules goes as follows. First, contracting parties wish to maximize the gains from trade and courts should embrace this goal in construing commercial agreements.⁴³² Second, private parties are better than judges at identifying efficient interpretive rules for their contractual relationships because they possess more information about transaction costs, enforcement costs, and the benefits of accurate interpretations.⁴³³ Third, “party preferences over interpretive rules are heterogeneous” because the optimal trade-off of costs and benefits varies from contract to contract.⁴³⁴ Given these three points, a court should defer to the parties’ choice of interpretive rules “just as it defers . . . to party preferences over a contract’s substantive terms.”⁴³⁵ In other words, when an agreement sets forth the interpretive approach that the parties wish the court to use in construing the agreement’s substantive provisions, the court should follow that instruction.⁴³⁶

5. *The Need for Empirical Evidence*

Most of the issues discussed in the prior four subparts raise questions of empirical fact that can only be answered with empirical evidence. The

⁴³¹ Schwartz & Scott, *supra* note 9, at 939 n.36 (“Recall that our theory holds only for the interpretation of contracts between business firms.”).

⁴³² *Id.* at 930.

⁴³³ *Id.* at 942, 944.

⁴³⁴ *Id.* at 930.

⁴³⁵ *Id.* at 930–31.

⁴³⁶ *Id.* at 942; *see also id.* at 943 (“Goal neutrality gives the parties control over the substantive terms of the contract. It takes an argument to reject the obvious implication that the parties should also have control over the rules that determine how those terms are identified and understood. . . . Sophisticated parties now can waive the right to a jury trial, or even the right to a trial in court, so they seemingly also should be able to waive the protection of exhaustive interpretive hearings.”). Schwartz and Scott further argue that the default interpretation rules ought to be textualist in nature because the majority of businesses favor that interpretive approach. *Id.* at 931, 940, 944–47, 955–57. For other sources arguing that contract interpretation doctrine should consist of default rules, see Bernstein, *Merchant Law*, *supra* note 413, at 1820–21 (contending that the Uniform Commercial Code should be amended to allow merchants to opt-out of either all or some of the Uniform Commercial Code’s contextualist provisions); Hermalin et al., *supra* note 9, at 90 (explaining that in the absence of certain special assumptions, and given that “it is difficult to draw strong general conclusions regarding how interpretation should proceed” because of the numerous ways that an interpretive regime can influence the parties, “perhaps the best that can be said is that private parties should be allowed the leeway to choose their favored interpretive regime—a leeway not always recognized by the legal system. . . .”).

bulk of the interpretation policy debate focuses on interpretive accuracy, transaction costs, and enforcement costs. Identifying which approach is most successful across these dimensions cannot be done in the abstract.⁴³⁷ But there are virtually no scholarly sources (or judicial opinions) that even purport to present systematic evidence on these questions.⁴³⁸ Of course, many of the theoretical, conceptual, and doctrinal arguments articulated in the caselaw and secondary literature are sophisticated and reasonably persuasive. But without quantitative empirical evidence and statistical analysis to support them, too many of these arguments merely constitute well-informed speculation. One goal of this article is to add to the small number of pieces, such as Lisa Bernstein's, that marshal genuine empirical evidence on fundamental questions of contract interpretation.

V. AN EMPIRICAL STUDY OF THE IMPACT OF THE TEXTUALIST AND CONTEXTUALIST APPROACHES ON ENFORCEMENT COSTS

A. *Study Outline*

Recall the policy issue my study is designed to address.⁴³⁹ Textualists contend that enforcement costs are lower under their approach to contract interpretation than under contextualism. Contextualists disagree. In fact, they argue that their system best minimizes litigation expenses. There are

⁴³⁷ See Cohen, *supra* note 272, at 148 (“The real question is which methodology has the lowest error rate and at what cost. It is hard to answer that question in the abstract.”); MITCHELL, *supra* note 316, at 102 (explaining that “commentators on formalism have *recognised* that neoformalism must be justified on the basis of empirical evidence”); *id.* at 114 (“Whether formalist or nonformalist judges will produce more errors depends on empirical evidence.”).

⁴³⁸ See MITCHELL, *supra* note 316, at 102 (noting that neoformalists/textualists “have not necessarily been concerned with providing [empirical] evidence” to support their claims”) (“One exception to this is Lisa Bernstein Other, law and economics, scholars have relied on ‘rational behavior models’ to make their point.”). As Mitchell observes, Lisa Bernstein’s empirical studies of adjudication practices in private, commercial courts, discussed *supra* at notes 413–17 and accompanying text, are a notable exception. See also Lawrence Solan et al., *False Consensus Bias in Contract Interpretation*, 108 COLUM. L. REV. 1268 (2008) (reporting the results of experimental studies conducted with both judges and laypeople regarding their interpretation of insurance contracts). For an example of an author writing on interpretation who properly acknowledges the weaknesses of his evidence, see Spigelman, *supra* note 328, at 429–30 (contending that businesses prefer textualist interpretation but noting that this claim is based on the author’s “own, necessarily limited experience” and that the author “knows of no empirical research” that supports the belief).

⁴³⁹ See *supra* text accompanying note 15; Part IV.B.3.

two basic components to enforcement costs—(1) the number of lawsuits filed, and (2) how long those lawsuits last. Each side maintains that it is superior on both elements; each asserts that fewer actions are commenced and that those actions end more quickly under its approach, reducing total enforcement costs.

Accordingly, there are actually two hypotheses that merit testing empirically rather than one: the textualist hypothesis and the contextualist counter-hypothesis. For each, the causal or independent variable is the school of interpretation employed by the courts and the dependent variables are the quantity of lawsuits filed and the length of those actions.⁴⁴⁰ The null hypothesis is that there is no difference between textualism and contextualism in their impacts on the number of contract interpretation lawsuits commenced and how long those cases last.

In a world of unlimited resources, I might test the two hypotheses through a comprehensive review of court files from contracts actions commenced across the country in both contextualist and textualist jurisdictions. Using such a method, I could directly measure the number of interpretation cases brought and the length of those actions. And I could attempt to control for variables such as contract complexity, the nature of the parties, and many of the important differences between states. But this is not a world of unlimited resources. And so I chose to use reported judicial opinions with certain key number classifications (the sample), from which I drew inferences about contract interpretation cases generally (the population).⁴⁴¹

⁴⁴⁰ See generally Epstein & King, *supra* note 7, at 35 (“All of these questions ask whether a particular ‘event’—the presence or absence of which we refer to as the *key causal variable* . . . caused a particular ‘outcome, or *dependent variable* . . .”).

⁴⁴¹ See generally *id.* at 31 (“It is up to the researchers, not the readers, to specify the target of their inference. Should that target be elusive or unclear to the investigators, what they might do is imagine how they would proceed with an unlimited budget and no limits on the amounts of time and effort they could expend.”); *id.* at 99–100 (“When we collect data to make inferences, a critical step is to identify the target population (or ‘population of interest’). This is all subjects, cases, countries, or other units in a specified time frame about which the researcher would collect information if time and resources were unconstrained.”) (emphasis removed).

B. Study Methodology, Raw Totals, and Statistical Analysis

To conduct the study, I began by identifying ten states where the caselaw appeared sufficiently clear to support including the jurisdiction rather firmly in either the textualist or contextualist camp. I sought five states for each category. And the ultimate classifications were based upon my own review of the doctrine in each jurisdiction.⁴⁴²

I also placed two additional restrictions on the selection of territories. First, each state needed to have an intermediate appellate court.⁴⁴³ Second, the two groups of states had to be of roughly comparable size in terms of population. The ten jurisdictions I settled on, with the population ranks set forth in parentheses, are listed in Table 2.⁴⁴⁴

Table 2. States Included in the Study

<i>Textualist</i>	<i>Contextualist</i>
Indiana (16)	Arizona (15)
Minnesota (21)	California (1)
Missouri (18)	Colorado (22)
New York (3)	New Jersey (11)
Texas (2)	Washington (13)

I turned next to which key numbers should be used in my Westlaw searches. There are thirty-five topics that fall under the general category of “Contracts,”⁴⁴⁵ and many of these topics contain key numbers that relate to interpretation. Examples include Sales, Insurance, Compromise and Settlement, and Release.⁴⁴⁶ In addition, numerous topics classified under

⁴⁴² The research supporting the textualist and contextualist classifications is on file with the author and will not be detailed here.

⁴⁴³ In conducting a key number study, one might focus solely on intermediate appellate courts. The advantage of such a design is that most appeals to these courts are of right whereas state supreme court jurisdiction is generally discretionary. Thus, focusing on intermediate appellate courts would eliminate at least one variable: differences in case selection by state high courts. However, there is an important weakness: typically, all state supreme court decisions are published while most intermediate appellate court rulings are not. *See supra* notes 155–58 and accompanying text. Which problem is more important—high court discretionary jurisdiction or intermediate court publication limitations—is unknown and may vary from project to project.

⁴⁴⁴ *See List of state and U.S. territories by population*, Wikipedia, http://en.wikipedia.org/wiki/List_of_U.S._states_and_territories_by_population (last visited Nov. 15, 2014).

⁴⁴⁵ WEST’S ANALYSIS OF AMERICAN LAW, *supra* note 69, at XI.

⁴⁴⁶ *Id.* at 301, 978–79, 1484, 1499–1500.

the other categories—Persons, Property, Torts, Remedies, and Government—implicate contracts.⁴⁴⁷ To illustrate, issues relating to damages for breach of contract are coded almost exclusively under the Damages topic, which is contained in the Remedies category.⁴⁴⁸ Principally because this was my first attempt at using the Key Number System as a coding device, I decided to focus on the core key numbers relating to contracts, contract interpretation, and the parol evidence rule, rather than construct a dataset using all potentially relevant key numbers.⁴⁴⁹

The label “Contracts” is used both for the third general digest category and for a distinct digest topic—“95. Contracts.”⁴⁵⁰ I concentrated on that topic and on “157. Evidence,” which falls under the Remedies category,⁴⁵¹ since one section of the Evidence key numbers concerns both the parol evidence rule and the use of extrinsic evidence in the construction of contracts and other writings.⁴⁵² I incorporated the key numbers that govern the parol evidence rule because of that rule’s extremely close relationship to contract interpretation. All other digest topics were excluded from the study.⁴⁵³

⁴⁴⁷ See generally *id.* at IX to XIV.

⁴⁴⁸ *Id.* at XII & 558–66.

⁴⁴⁹ Another reason I settled on a narrower approach is that a broader study involves additional technical and methodological issues that I have not yet resolved. For example, the character limitation on Westlaw searches might make it impossible to run the queries necessary to conduct a truly comprehensive key number study of contract interpretation using all of the pertinent digest topics. See *supra* notes 268–69 and accompanying text (discussing the character limit governing Westlaw searches). I spent some time working with West reference attorneys to solve this problem, but we were unable to find an answer. However, I have not finished investigating the issue. If I am able to resolve the problem—and a few others—I plan to conduct a broader study and publish (1) the substantive results, (2) the solution to the technical problem created by the character limitation, and (3) the solutions to various other difficulties that may arise with more comprehensive key number projects.

⁴⁵⁰ *Id.* at XI & 370–81.

⁴⁵¹ *Id.* at XII, XVI.

⁴⁵² *Id.* at 706–10 (“XI. Parol or Extrinsic Evidence Affecting Writings.”).

⁴⁵³ A brief note is in order regarding the key numbers that concern Article 2 of the Uniform Commercial Code (topic 343—“Sales”), a statute that is at the center of contract law. Article 2 embraces contextualist interpretation. See U.C.C. § 1-303(d); *id.* § 2-202 & cmt. 1(c). Thus, courts in every state *should* use contextualism when construing agreements governed by the Code. If the courts are complying with the law, then cases tagged with Sales key numbers are irrelevant to my study because there are no measurable legal differences between the states when it comes to contracts for the sale of goods. Now, at least one respected source contends that judges frequently ignore the dictates of Article 2 and use a form of textualism, requiring a finding of facial ambiguity before admitting extrinsic evidence for interpretive purposes. See WHITE ET AL., *supra* note 316, § 3:14, at 221–22. But this was not enough to persuade me that Sales key numbers ought to be included in the study. For a discussion of a type of

I then analyzed the various interpretation and parol evidence rule key numbers under topics 95 (Contracts) and 157 (Evidence). This involved running more than one hundred searches in the ALLCASES database on Westlaw to identify precisely which key numbers should be used in my ultimate queries.⁴⁵⁴ To illustrate, these searches lead to the exclusion of the key number that deals primarily with the duty of good faith,⁴⁵⁵ which is generally considered a separate topic from interpretation and the parol evidence rule.

Next, I organized the selected key numbers into three groups: (1) contract interpretation key numbers from the Contracts digest topic, (2) contract interpretation key numbers from the Evidence digest topic, and (3) parol evidence rule key numbers from the Evidence topic. Using these groups, I constructed three search queries for each state. The first query contained only the key numbers in group (1). The second contained the key numbers in groups (1) and (2). And the third contained the key numbers in all three groups—(1), (2), and (3). The three queries were run in both the relevant state database and the relevant mixed state/federal database, so there were actually six searches in total for each jurisdiction. I included the mixed state/federal databases in the study because many cases applying state contract law are litigated in federal court under diversity jurisdiction.⁴⁵⁶ The results of the six searches served as proxies for the level of contract *interpretation* litigation in each state during the relevant time period.⁴⁵⁷

project where Sales key numbers *would* have a role to play, see *infra* notes 514–17 and accompanying text. Note also that based on several tests I ran, cases applying the U.C.C. rather than the common law probably did not create any bias in the dataset I created employing only the Contracts and Evidence key numbers.

⁴⁵⁴ Such a search would now be run on WestlawNext with “All States” and “All Federal” selected in the dropdown menu next to the search box on the home page.

⁴⁵⁵ 95k168. “Terms implied as part of contract.”

⁴⁵⁶ To limit the cases to the appropriate jurisdiction in the mixed state/federal databases, additional limiting search language was necessary. The precise searches I used are set forth in Appendix I. Note also that the dataset that includes federal decisions is arguably superior because it includes a much higher number of trial court opinions. However, because the federal trial court opinions included in the digest are such a small percentage of the total number of trial court orders and lawsuits, see *supra* notes 152–53 and accompanying text, the inclusion of federal trial court decisions might do more to skew the sample set than improve it.

⁴⁵⁷ After I began work on this article, Westlaw was replaced by WestlawNext. However, I reran all of my searches in WestlawNext, and the raw number of cases recovered by each search changed by,

At this stage, there was still one piece missing. I could not simply compare the search results from the textualist states to those from the contextualist states. That is because there are far too many differences between the ten jurisdictions included in my study to make such a comparison fruitful—differences in population, business activity, the number of contracts executed per person, litigation culture, rules of procedures, opinion publishing practices,⁴⁵⁸ and countless others.⁴⁵⁹ And using states with comparable populations and intermediate appellate courts⁴⁶⁰ barely scratches the surface of what would be necessary to account for these variations. Therefore, I needed to add something to the study that attempts to control for the differences among my ten states.⁴⁶¹

The control I decided to employ was a search in every relevant Westlaw database for all cases that use *any* key number from the Contracts digest topic (again, topic 95). This search served as a proxy for the total level of *contract* litigation in each state during the time period of the study. With the addition of this query, I could compare the cases coded for contract interpretation to the cases coded as contract disputes generally. If textualism and contextualism—in relation to each other—do not impact how many lawsuits are filed or how long those lawsuits last, then one might reasonably expect the ratio of contract interpretation cases to general contract cases to be roughly the same in textualist and contextualist states. Put another way, if the ratio of cases returned by my three queries to the

at most, a trivial amount. For example, all of the New York searches for state court cases returned identical results in Westlaw and WestlawNext. And the New York searches for state and federal cases in WestlawNext returned one additional case for queries one, two, and three, and six additional cases for the control query (discussed in the next two paragraphs). Likewise, every Texas search was either identical in Westlaw and WestlawNext or differed by a single case.

⁴⁵⁸ For example, opinions from trial courts in New York (a textualist state) and New Jersey (a contextualist state) are reported sufficiently often to warrant citation rules for that type of case. See THE BLUEBOOK, *supra* note 151, at 277, 280. The other eight states in my dataset do not publish enough trial court decisions to justify such rules. See *id.* at 250, 252, 254, 263, 272, 274, 295, 300.

⁴⁵⁹ See Korobkin, *supra* note 4, at 1063 (“Studying the effects on contracting behavior of specific contract rules by comparing across jurisdictions or time periods raises a number of methodological problems, most prominent among them how to account for the many other known and unknown variables, besides the differences in contract law, that could also account for differences in observed behavior.”); see also Epstein & King, *supra* note 7, at 44 (“The fact is that even the best empirical research can be inadvertently affected by hundreds of confounding factors.”).

⁴⁶⁰ See *supra* notes 443–44 and accompanying text.

⁴⁶¹ See Epstein & King, *supra* note 7, at 78 (identifying the types of confounding variables that researchers should attempt to control for).

cases returned by the control search is constant for textualist and contextualist states, then the null hypothesis cannot be rejected. Alternatively, if the fraction of contracts cases that are interpretation cases is higher in contextualist jurisdictions by a statistically significant amount, this suggests that (i) parties file more interpretation lawsuits in contextualist states, (ii) the interpretation lawsuits in those states last longer, as demonstrated by the fact that more reach the appellate level or otherwise result in a reported decision,⁴⁶² or (iii) some combination of the two. And the reverse is true if the fraction is higher in textualist states by a statistically significant level.⁴⁶³

I chose cases classified with a key number from topic 95 as the control for three reasons. First, I wanted the study to rely as much as possible on key number searches. Second, a control that focuses on reported decisions reduces the likelihood that any differences in opinion publishing practices across states will bias my results. That is because such differences should apply equally to contract interpretation cases and general contract cases.⁴⁶⁴ Third, numerous other factors that vary among the states in my study—such as population, contracts executed per person, and rules of procedure—probably impact the levels of contract interpretation litigation and general contract litigation in many of the same ways. Thus, comparing interpretation cases specifically to contract cases generally should control for a large number of differences between the relevant states.⁴⁶⁵ Of course, the control is far from perfect (as discussed more in the next subpart).⁴⁶⁶

⁴⁶² Recall that most reported cases with key numbers are appellate cases. *See supra* note 150 and accompanying text.

⁴⁶³ Of course, all of this assumes that reported decisions with key numbers are representative of litigation generally. That is likely not the case as explained previously. *See supra* notes 168–205 and accompanying text; *see also infra* note 484 (addressing the application of inferential statistics to a non-random sample).

⁴⁶⁴ The first two reasons together explain why I did not compare reported decisions tagged with interpretation key numbers in a given state to all civil filings in that state classified as contract disputes by the local administrative office of the courts (assuming such information is available in every relevant territory).

⁴⁶⁵ Comparing contract interpretation litigation directly to measures like population, contracts executed per person, state gross domestic product, total civil filings, or even an index of these types of factors, will likely control for far fewer differences across states. With that said, such comparisons may provide helpful information and thus might be worth pursuing in future work.

⁴⁶⁶ *See infra* notes 498–503 and accompanying text.

Nonetheless, I think it is the best candidate given the purposes of this project.

Note also that in measuring the impacts of the two interpretive approaches, my study protocol does not distinguish between the two aspects of enforcement costs—the number of lawsuits filed and how long those lawsuits last. As explained above, any causal effect that is suggested by an analysis of my key number searches could flow from a change to either parameter or a combination of the two.⁴⁶⁷ However, this is not a problem. The debate between textualists and contextualists is focused on overall enforcement costs. Accordingly, any results that bear upon that issue are valuable even if the precise mechanism through which interpretive approach increases enforcement costs is unidentifiable.

Five additional concerns regarding my study methodology are worth discussing. First, recall the problem of “under-coding” explained previously in the section regarding the advantages and disadvantages of using key numbers in empirical research.⁴⁶⁸ To recap, there is significant overlap among many of the digest topics and subtopics. As a result, headnotes that could have been classified under a given topic are frequently classified instead under a different, related topic. This raises an issue for my study because I used only (1) the interpretation key numbers in topics 95 (Contracts) and 157 (Evidence) to gather interpretation decisions, and (2) topic 95 to gather contract decisions. Accordingly, I excluded numerous cases that involve interpretation specifically or contracts generally. For example, during the time period of my study, state courts in the ten relevant jurisdictions issued 823 opinions that are coded with contract interpretation key numbers from the Insurance topic, but that contain none of the Contracts and Evidence key numbers I used to collect interpretation decisions.⁴⁶⁹

⁴⁶⁷ See *supra* text accompanying notes 462–63. To repeat, if the ratio of interpretation cases to general contract cases recovered by my searches is higher in one set of states, this suggests that (i) more interpretation actions are commenced in that set of jurisdictions, (ii) interpretation lawsuits last longer in that set, or (iii) a combination of each.

⁴⁶⁸ See *supra* notes 246–52 and accompanying text.

⁴⁶⁹ To see this, run the following search on Westlaw for all state cases: co(in) co(mn) co(mo) co(ny) co(tx) co(az) co(ca) co(co) co(nj) co(wa) & 217XIII(G) & DA(aft 12/31/1999) & DA(bef 01/01/2010) % (157XI(D) (95k143 95k143.5 95k147 95k148 95k149 95k150 95k151 95k160 95k161 95k162 95k163 95k164 95k165 95k166 95k167 95k169 95k170 95k171 95k172 95k173 95k174 95k175 95k176) (157XI(A) 157k439 157k440 157k441 157k442 157k443 157k444)).

It is unclear how under-coding impacted my findings. Indeed, determining the effect would probably require running a full study using more of the relevant key numbers.⁴⁷⁰ For now, the important point is this: It is possible that a methodology employing a larger group of key numbers would lead to a different result.⁴⁷¹

Second, my key number searches retrieved cases for the specific interpretation datasets that do not actually concern contract interpretation. The searches were over-inclusive for two reasons. First, key numbers are often added to opinions even though the associated topic was not litigated. I previously labelled this the problem of “over-coding.”⁴⁷² Second, several of the key numbers that address contract interpretation suffer from overbreadth—they also cover the construction of other types of legal documents, such as wills.⁴⁷³ When I discussed these limitations above, I hypothesized that for most empirical studies using large samples, any over-coding or overbreadth is likely to balance out. But I also conceded that each limit might pose genuine problems for certain projects.⁴⁷⁴

Accordingly, I conducted several tests for the prevalence of over-coding and overbreadth within my sets of interpretation cases. These tests included manually reviewing a quasi-random sample of 400 cases in the largest set (query 3, state and federal)—200 from textualist states and 200 from contextualist states—and running dozens of searches on Westlaw, both generally and within my dataset, for key numbers and terms that signal

⁴⁷⁰ For additional information on the possibility of conducting a broader study, see *supra* note 449; *infra* text accompanying note 511.

⁴⁷¹ To elaborate, recall that my study excluded 823 state court cases coded with key numbers from the Insurance digest topic that concern the interpretation of insurance agreements. See *supra* text accompanying note 469. Of the cases, 482 were from the five textualist states and 341 were from the five contextualist states. If those cases were added to the results for query 3 as run in the relevant state court databases, the ratio of interpretation cases to contract cases would change significantly. As the reader will see shortly, the ratio of interpretation cases to contract cases in state courts was 5.0% higher in *textualist* states using the query 3, weighted measure. But if the insurance cases were included, the ratio would have been 7.0% higher in *contextualist* states. See *infra* Table 3. That is a considerable shift. Of course, Insurance is only one of several digest topics excluded by my study protocol. But this example should illustrate how the under-coding problem might, in theory, have dramatically impacted my findings.

⁴⁷² See *supra* notes 241–44 and accompanying text. Note that the control dataset includes opinions that do not involve contract litigation for the same reason.

⁴⁷³ See *supra* notes 257–59 and accompanying text.

⁴⁷⁴ See *supra* notes 244, 259 and accompanying text.

whether an opinion suffers from over-coding or overbreadth. While the various tests I employed did not yield identical results, the outcome of the most comprehensive test suggests that over-coding and overbreadth probably do balance out in my dataset.⁴⁷⁵ Nonetheless, given the inconsistency in my test findings, it is possible that over-coding and overbreadth distorted my results.⁴⁷⁶

Third, my study addresses how each state's approach to construing contracts impacts the level of interpretation litigation in that jurisdiction. However, courts located in one state often apply the law of another state or federal law. If a significant number of cases in either the textualist set of interpretation cases or the contextualist set (but not both) apply the law of a foreign jurisdiction, then the study is no longer comparing apples to apples.

As with the problems of over-coding and overbreadth, I used a variety of techniques to test for the incidence of cases applying foreign law in my dataset. This included reviewing the same 400 decisions noted two paragraphs above and running multiple searches on Westlaw. These tests support two conclusions. First, virtually every interpretation case applying foreign law in my dataset was filed in federal court. The study findings based solely on state court decisions were thus almost certainly not biased by the incidence of foreign law opinions. Second, substantially more federal cases from the contextualist states concerned foreign law than from the textualist states. Therefore, the study results based on both state *and* federal decisions may have been biased by foreign law cases.⁴⁷⁷

Fourth, there can be more than one reported opinion from the same lawsuit, such as where an appellate decision is appealed to the state

⁴⁷⁵ To elaborate, of the cases in the largest interpretation set (again, query 3, state and federal), 18.84% of the textualist decisions and 18.42% of the contextualist decisions contain markers indicating they likely were included as a result of over-coding or overbreadth. Those results comes close to perfectly balancing out. The precise protocol used to obtain these figures is on file with the author.

⁴⁷⁶ Given that under-coding, over-coding, and overbreadth all have significant potential to corrupt the integrity of datasets created using contract interpretation key numbers, it is possible that contract interpretation is not the best subject for empirical work employing the American Digest System. Further research will be necessary to assess this theory.

⁴⁷⁷ For example, one of the tests I conducted on the largest set (query 3, state and federal) found that 304 out of 2548 textualist decisions, or 11.93%, and 189 out of 847 contextualist decisions, or 22.31%, contained markers indicating they likely applied foreign law. This result supports the conclusion that the foreign law decisions do not balance out, unlike the over-coded and overbroad cases. Additional details regarding this test are on file with the author.

supreme court.⁴⁷⁸ However, based on my review of the interpretation and control groups, there are simply too few cases that (1) are from the same lawsuit as another published opinion, and (2) contain the same pertinent key numbers as the related decision, to corrupt the integrity of my data.

Fifth, when running searches for state and federal cases together, I have found West's search algorithm to be somewhat less reliable than normal. In particular, not all federal decisions applying the law of a state are retrieved in such searches.⁴⁷⁹ And sometimes cases are found by these searches that have nothing to do with the relevant state.⁴⁸⁰ Unfortunately, I am not sure how common this problem is. Determining the scope will require additional analysis in subsequent work.

There are two final aspects of my methodology worth noting. First, I relied solely on the Key Number System for the data coding in this study. I did not exclude from or include in the dataset a single case based on my own analysis of the opinion. And I read only a small fraction of the cases in my sample.⁴⁸¹ Second, all searches were run with a date restriction that limited the dataset to cases decided from January 1, 2000, through December 31, 2009.

The next two pages set forth the raw totals from my study. The first page is the data that is exclusively from state courts. The second page contains the data from state and federal courts combined.

⁴⁷⁸ See Zirkel & Johnson, *supra* note 114, at 7 (observing that using key number searches to measure the quantity of education-related litigation results in some degree of double counting because there are cases "that have multiple different decisions due to appeals, remands, and separable issues extending from the threshold stage, such as discovery or statutes of limitations, to the post-trial stage, such as attorney's fees"); Zirkel & Richardson, *supra* note 109, at 769 (same).

⁴⁷⁹ For example, my search for the relevant key numbers in Arizona caselaw—both state and federal—did not find *GTE Wireless, Inc. v. Cellexis Int'l, Inc.*, 341 F.3d 1 (1st Cir. 2003), even though this decision contained one of the relevant key numbers and applied Arizona law, *id.* at 4–5.

⁴⁸⁰ For example, my search for Arizona state and federal authorities included a case from the Supreme Court of Pennsylvania decided in 1792—*Field for the use of Oxley v. Biddle*, 2 Dall. 171 (Pa. 1792)—which is nearly seventy years before the beginning of the Arizona Reports. Likewise, my search for Illinois state and federal authorities retrieved *Utica Mutual Insurance Co. v. Vigo Coal Co.*, 393 F.3d 707 (7th Cir. 2004), which clearly applies Indiana law, *id.* at 715–16, and is an appeal in a lawsuit that originated in the Federal District Court for the Southern District of Indiana, *id.* at 707. (Note that Illinois was not included in my study.)

⁴⁸¹ For examples of reasons why I read some of the cases, see *supra* notes 472–77, 479–80 and accompanying text.

Table 3: State Court Cases Only (2000-2009)

	Query 1	Query 2	Query 3	Control Search
	Group 1 95(II)(A) Selected ⁴⁸²	Group 1 95(II)(A) Selected + Group 2 157XI(D) All	Group 1 95(II)(A) Selected + Group 2 157XI(D) All + Group 3 157XI(A) All & 157XI(C) Selected	95 All
<u>Textualist</u>				
Indiana	229–53.9%	240–56.5%	251–59.1%	425
Minnesota	64–44.8%	70–48.9%	71–49.7%	143
Missouri	219–48.0%	236–51.8%	243–53.3%	456
New York	382–31.7%	449–37.2%	497–41.1%	1208
Texas	566–45.5%	610–49.0%	645–51.8%	1244
<i>Total–Weighted*</i>	1460–42.0%	1605–46.2%	1707–49.1%	3476
<i>Total–Unweighted**</i>	44.8%	48.7%	51.0%	
<u>Contextualist</u>				
Arizona	29–33.0%	32–36.4%	33–37.5%	88
California	177–30.4%	223–38.3%	249–42.7%	583
Colorado	84–48.9%	89–51.7%	91–52.9%	172
New Jersey	74–35.9%	84–40.8%	86–41.7%	206
Washington	96–38.7%	111–44.8%	113–45.6%	248
<i>Total–Weighted</i>	460–35.5%	539–41.6%	572–44.1%	1297
<i>Total–Unweighted</i>	37.4%	42.4%	44.1%	
<u>Textualist minus Contextualist</u>				
<i>Weighted Unweighted</i>	6.5%	4.6%	5.0%	
<i>Unweighted</i>	7.4%	6.3%	6.9%	

***“Weighted” means the sum of the interpretation *cases* from all five states divided by the sum of all contracts *cases* from the same five states. Larger states have more impact under this measure.

***“Unweighted” means the sum of the *percentages* in each column divided by five. Each state has the same impact under this measure.

⁴⁸² Included with each group is a brief description of the key numbers that constitute that group using terminology from West’s outline of the topic. The precise key numbers that constitute each group are identified in Appendix 1.

Table 4: State and Federal Court Cases (2000-2009)

	Query 1	Query 2	Query 3	Control Search
	Group 1 95(II)(A) Selected	Group 1 95(II)(A) Selected + Group 2 157XI(D) All	Group 1 95(II)(A) Selected + Group 2 157XI(D) All + Group 3 157XI(A) All & 157XI(C) Selected	95 All
<u>Textualist</u>				
Indiana	280-50.5%	296-53.4%	307-55.6%	554
Minnesota	119-40.3%	134-45.4%	145-49.2%	295
Missouri	266-45.3%	289-49.2%	301-51.3%	587
New York	761-32.2%	875-37.1%	961-40.7%	2361
Texas	732-44.7%	789-48.1%	834-50.9%	1639
<i>Total-Weighted</i>	<i>2158-39.7%</i>	<i>2383-43.8%</i>	<i>2548-46.9%</i>	<i>5436</i>
<i>Total-Unweighted</i>	<i>42.6%</i>	<i>46.6%</i>	<i>49.5%</i>	
<u>Contextualist</u>				
Arizona	45-28.5%	49-31.0%	51-32.3%	158
California	287-28.2%	359-35.3%	403-39.6%	1017
Colorado	112-43.1%	120-46.2%	124-47.7%	260
New Jersey	108-29.9%	123-34.1%	128-35.5%	361
Washington	116-37.2%	138-44.2%	141-45.2%	312
<i>Total-Weighted</i>	<i>668-31.7%</i>	<i>789-37.4%</i>	<i>847-40.2%</i>	<i>2108</i>
<i>Total-Unweighted</i>	<i>33.4%</i>	<i>38.2%</i>	<i>40.1%</i>	
<u>Textualist minus Contextualist</u>				
<i>Weighted</i>	<i>8.0%</i>	<i>6.4%</i>	<i>6.7%</i>	
<i>Unweighted</i>	<i>9.2%</i>	<i>8.4%</i>	<i>9.4%</i>	

Some additional explanation of the data is in order. First, the percentages listed for each state reflect the ratio of cases recovered by queries 1, 2, and 3, to the cases recovered by the control search. For example, in Table 3, query 1 returned 280 cases in Indiana. The control search for that state returned 425 cases. Two-hundred eighty is 53.9 percent of 425. Second, I combined the numbers for the five textualist states and for the five contextualist states in two ways—a weighted total and an unweighted total. To illustrate, in Table 3, the combined *weighted* total for the textualist states under query 1 is 42.0%. That figure was created by adding up all of the query 1 results for each state (1460) and dividing that number by the sum of all the control search results for those states (3476). Using this measure, the states with more cases have greater impact on the ultimate percentage. By comparison, the combined *unweighted* total for the textualist states under query 1 is 44.8%. That number was created by adding all of the query 1 percentages (53.9%, 44.8%, 48.0%, 31.7%, and 45.5%) and dividing by five. Using this measure, each state has equal impact on the ultimate percentage.

Turning to analysis of the data, and focusing first on the raw numbers, states using a textualist approach had noticeably more contract interpretation litigation as a fraction of general contract litigation than states using a contextualist approach. Most relevant are the weighted and unweighted combined totals for each group of five states. Consider the unweighted figures in Table 3. For the ten years of the study, interpretation litigation constituted 44.8%, 48.7%, and 51.0% of total contract litigation in textualist state courts, depending on the measure used for interpretation cases (queries 1, 2, and 3). For contextualist states, the numbers are 37.4%, 42.4%, and 44.1%. The smallest difference between textualist and contextualist states on these measures is 6.3% (query 2).

Table 5 lists all of the differences between the combined textualist percentages and the combined contextualist percentages from the bottom of Tables 3 and 4.

Table 5: Differences Between Combined Totals from Tables 3 and 4

	Table 3: State Court Cases Only		
	Query 1	Query 2	Query 3
<u>Textualist minus Contextualist</u>			
<i>Weighted</i>	6.5%	4.6%	5.0%
<i>Unweighted</i>	7.4%	6.3%	6.9%

	Table 4: State and Federal Court Cases		
	Query 1	Query 2	Query 3
<u>Textualist minus Contextualist</u>			
<i>Weighted</i>	8.0%	6.4%	6.7%
<i>Unweighted</i>	9.2%	8.4%	9.4%

As this table makes clear, the ratio of contract interpretation litigation to general contract litigation was larger in the textualist states under all twelve measures used. The greatest difference between textualist and contextualist states was 9.4% (query 3, unweighted, state and federal). The smallest was 4.6% (query 2, weighted, state only).

While the raw totals are somewhat illuminating, they lack genuine persuasive force standing alone. Accordingly, I retained professional statisticians from the University of Georgia Statistical Consulting Center⁴⁸³ to analyze the data.⁴⁸⁴ The statisticians ran t-tests on each of the twelve

⁴⁸³ See *Statistical Consulting Center*, UNIV. OF GA. FRANKLIN COLL. OF ARTS AND SCI. (last visited Feb. 27, 2016), <http://www.stat.uga.edu/consulting>. The center performs statistical analysis for academics who lack the training to conduct such analysis themselves. They were extremely helpful in writing this article. My primary contact was Associate Director Kimberly Love-Meyers (who has since moved on to another job).

⁴⁸⁴ Technically, the use of inferential statistics is valid only when applied to a random sample. Konrad Lajer, *Statistical Tests as Inappropriate Tools for Data Analysis Performed on Non-Random Samples of Plant Communities*, 42 *FOLIA GEOBOTANICA* 115, 116–18 (2007) (“Standard statistical tests cannot be validly used with non-random samples. . . .”); Phillips et al., *supra* note 232, at 89 (“In opposition to probability samples, nonprobability samples tend to be biased and are not generalizable, and so are not recommended for statistical analysis.”); Sarah H. Ramsey & Robert F. Kelly, *Social Science Knowledge in Family Law Cases: Judicial Gate-Keeping in the Daubert Era*, 59 *U. MIAMI L. REV.* 1, 68 (2004) (“First, if the study’s sample was not selected randomly . . . it is inappropriate to report the level of statistical significance for relationships among variables found in the sample. . . . Second, a corollary to the first point is that the appropriate use of tests of statistical significance has

measures and logistic regressions on two measures—(1) query 1, weighted, state courts, and (2) query 1, weighted, state and federal courts.⁴⁸⁵ The two logistic regressions found no statistically significant difference between the contextualist and textualist jurisdictions. And the t-tests also found no statistically significant difference for eleven of twelve measures. On the twelfth measure—query 3, unweighted, state and federal—the higher level of litigation in textualist states was statistically significant at the .05 level. This measure had the largest raw difference between textualist and contextualist states—9.4%.⁴⁸⁶ And it is the most comprehensive; it included all the relevant key numbers and covered both state and federal courts. The twelfth measure thus used the biggest dataset.⁴⁸⁷ However, given that eleven of twelve t-tests and the two logistic regressions failed to find a statistically significant difference, the ultimate conclusion of this study is that the null hypothesis cannot be rejected: The study cannot reject the

everything to do with whether the sample was randomly selected Studies using large nonrandom samples have no more legitimate call to use tests of statistical significance than small convenience samples.”); *id.* at 68–69 n.237 (“What is key here is that the use of confidence levels/intervals involves the same statistical reasoning as significance tests and that each requires a randomly drawn sample in order to be validly employed.”); *see also* Kevin S. Marshall et al., *The Habit Evidence Rule and its Misguided Judicial Legacy: A Statistical and Psychological Primer*, 36 LAW & PSYCHOL. REV. 1, 31–32 & n.200 (2012) (“Inferential statistics is a body of scientific and mathematical methods used to draw conclusions or inferences about unknown characteristics of a population on the basis of known sample data.”) (collecting authorities); Brian Root, *Statistics and Data in Human Rights Research*, 107 AM. SOC’Y INT’L L. PROC. 65, 66 (2013) (“Generally speaking, all statistics fall into two broad categories: descriptive and inferential. Descriptive statistics simply describe the data that were collected. Inferential statistics use observed data to predict what is true of areas beyond the data.”). This creates a problem for those who wish to use statistical analysis to draw inferences from a sample created using the Key Number System since such a sample is seldom a random subset of the population under study. *See supra* notes 168–205 and accompanying text; *see also infra* text accompanying notes 536–37 (describing a second sampling problem with the specific protocol followed in this study). However, when the choice is between (1) relying solely on the raw totals of a non-random sample (i.e., relying solely on descriptive statistics), and (2) relying on both the raw totals *and* statistical analysis of the non-random dataset (i.e., relying on both descriptive and inferential statistics), the latter is the superior approach. Indeed, inferential statistics are regularly applied to non-random datasets. *See, e.g.*, Georgakopoulos, *supra* note 115, at 127–30; McChesney, *supra* note 56, at 170–85; *see also* Hoffman et al., *supra* note 153, at 687 (explaining that scholars who analyze the legal system by using samples of judicial opinions normally “claim that that the dataset of opinions is good enough for statistical inference”). This should be unsurprising given that the use of non-random samples is justified in many contexts. *See supra* note 231.

⁴⁸⁵ Appendix 2 explains why there are only two logistic regressions.

⁴⁸⁶ *See supra* Table 5.

⁴⁸⁷ Technically, measure 12 was tied for largest. Measure 9—query 3, weighted, state and federal—used the same dataset as measure 12.

hypothesis that there is no difference between textualism and contextualism in their impacts on the number of contract interpretation lawsuits filed and the length of those proceedings. Put another way, the study finds no support for the textualist hypothesis that contextualism has higher enforcement costs and very little support for the contextualist counter-hypothesis that textualism has higher enforcement costs.⁴⁸⁸

C. Discussion

This subpart discusses potential explanations for my failure to find a statistically significant difference in enforcement cost levels between textualism and contextualism. The explanations are of two types. Some reject the finding as invalid because of problems with my study methodology. Others presume that the result is valid and constitute theories as to why the two approaches to contract interpretation might have similar impacts on enforcement costs.

1. Methodological Explanations for the Study Findings

There are multiple grounds upon which one could challenge the results of this study. First, recall the many qualifications regarding my protocol discussed above. Most importantly, the study attempts to draw inferences about the population of all cases from a sample of reported cases.⁴⁸⁹ But the latter is probably not a representative subset of the former. That is a considerable weakness.⁴⁹⁰ Also critical is the fact that there are countless differences between the ten states in the study that I did not—and probably could not—control for.⁴⁹¹ The other problems noted previously,⁴⁹² such as concerns with West's coding protocols, are not as serious as the first two issues. But taken together, they likewise justify healthy skepticism about the study results.

⁴⁸⁸ For additional details regarding the statistical tests, see Appendix 2.

⁴⁸⁹ See *supra* Parts V.A.–B.

⁴⁹⁰ See *supra* notes 168–205 and accompanying text; see also *supra* note 484.

⁴⁹¹ See *supra* notes 458–65 and accompanying text.

⁴⁹² See *supra* notes 235–69, 467–80 and accompanying text.

Second, remember that the caselaw regarding contract interpretation is extraordinarily convoluted.⁴⁹³ This complexity sometimes makes it difficult to classify a state as textualist or contextualist.⁴⁹⁴ Indeed, most jurisdictions fall somewhere along the continuum between the two approaches rather than clearly on one side. While I believe I selected ten states for the study that fit firmly in either the textualist or contextualist camp, my classification decisions are not immune from criticism. For example, I categorized New Jersey and Arizona as contextualist. But there are multiple precedents in each jurisdiction that apply the textualist approach.⁴⁹⁵ Similarly, Texas courts consistently note that textualism is the governing system in their state.⁴⁹⁶ Yet they sometimes implement that approach in a manner that looks more like contextualism.⁴⁹⁷ Given these examples, one might object that the groups of states I chose are too similar in their contract interpretation practices for any differences between textualism and contextualism to show up in the study. An experiment with territories that employ purer forms of the two approaches could lead to different results. Unfortunately, such a project may not be possible because there are likely too few states that use textualist and contextualist interpretation in their ideal forms.

⁴⁹³ See *supra* notes 312–16 and accompanying text.

⁴⁹⁴ See *supra* note 316 and accompanying text.

⁴⁹⁵ Regarding New Jersey, compare *Conway v. 287 Corporate Center Associates*, 901 A.2d 341, 346–47 (N.J. 2006) (explaining that New Jersey has adopted contextualist contract interpretation), with, for example, *Barr v. Barr*, 11 A.3d 875, 882 (N.J. Super. Ct. App. Div. 2010) (stating that extrinsic evidence may only be considered in interpreting a contract if the language on the face of the agreement is reasonably susceptible to more than one meaning, which is the textualist approach). Regarding Arizona, compare *Taylor v. State Farm Mutual Automobile Insurance Co.*, 854 P.2d 1134, 1138–41 (Ariz. 1993) (explaining that Arizona has adopted the contextualist approach), with, for example, *Scalia v. Green*, 271 P.2d 479, 482 (2011) (“In interpreting an easement created by deed or grant, we apply the rules of contract construction. . . . When a deed is unambiguous, we will not consider extrinsic evidence of the parties’ intent.”).

⁴⁹⁶ See, e.g., *David J. Sacks, P.C. v. Haden*, 266 S.W.3d 447, 450–51 (Tex. 2008) (“An unambiguous contract will be enforced as written, and parol evidence will not be received for the purpose of creating an ambiguity. . . . Only where a contract is ambiguous may a court consider the parties’ interpretation and admit extraneous evidence to determine the true meaning of the instrument.”) (internal quotation marks omitted) (citations omitted).

⁴⁹⁷ See LINZER, 6 CORBIN ON CONTRACTS, *supra* note 310, § 25.14[a], at 158 (further arguing that the Texas Supreme Court’s interpretive method “misleads planners into thinking they can rely on plain meaning when in fact the courts are not that rigid”); *id.* at 155–61.

Third, a textualist could object that my study failed to take into account that the interpretive approach a state employs is generally correlated with other rules of contract law in that jurisdiction. Textualism is a “classical” doctrine. States that use that method of interpretation also typically follow other classical doctrines. Likewise, contextualism is a “modern” doctrine. Jurisdictions that employ this approach normally conform to modern contract principles.⁴⁹⁸ Classical contract law is marked by clear rules and strict adherence to legal formalities such as the statute of frauds. By contrast, modern contract law favors general standards, such as “good faith” and “unconscionability,” and shows greater sympathy for equitable precepts.⁴⁹⁹ Many scholars, especially those in the field of law and economics, contend that legal norms that take the form of general standards cause more litigation than narrow rules.⁵⁰⁰ This creates a potential problem for my study protocol.

To measure the impact of interpretive approach on enforcement costs, I compared the ratios of contract interpretation litigation to general contract litigation in textualist and contextualist states—with the levels of each type

⁴⁹⁸ See Miller, *supra* note 414 (discussing numerous differences between California’s “contextualist” contract law and New York’s “formalist” contract law, including differences regarding formation, defenses, interpretation, choice-of-law clauses, and arbitration clauses). See also Grumman Allied Indus., Inc. v. Rohr Indus., Inc., 748 F.2d 729, 733–34 (2d Cir. 1984) (referring to the textualist and contextualist approaches to contract interpretation as the “classical” and “modern” approaches).

⁴⁹⁹ See KNAPP ET AL., *supra* note 273, at 31–32.

⁵⁰⁰ See Ehud Guttel & Alon Harel, *Uncertainty Revisited: Legal Prediction and Legal Postdiction*, 107 MICH. L. REV. 467, 481 & nn.84, 85 (2008) (“The use of rules and standards, as law and economics scholars have shown, involves different costs and benefits. . . . [S]tandards usually entail higher enforcement and compliance costs than rules.”) (collecting authorities); Russell B. Korobkin, *Behavioral Analysis and Legal Form: Rules vs. Standards Revisited*, 79 OR. L. REV. 23, 56 (2000) (concluding that combining behavioral analysis with economic analysis leads to the insight that “standards will be more expensive to apply [than rules] both because applying a standard usually will be more expensive than applying a rule and because more cases will be litigated in a standards regime than in a rules regime. . . .”); Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557, 562–63, 570, 572 (1992) (explaining why rules should have lower enforcement costs than standards); *but see id.* at 573 n.35, 575 n.41 (discussing various reasons why there might be more or less litigation under standards than under rules). There are two principal bases for this conclusion. First, standards increase litigation because they generate more uncertainty than rules. See Korobkin, *supra*, at 32 (“The ex ante certainty that rules provide should encourage more disputes to settle out of court and not require adjudication at all.”). Second, when employing rules, adjudicators need merely decide whether the relevant facts are present. *Id.* But under standards, adjudicators must both uncover the facts and determine the precise content of the law since this was not done by the enactor *ex ante*. *Id.*; accord Kaplow, *supra*, at 562–63, 570. The extra step requires the expenditure of additional resources.

of litigation being determined by the results of my key number searches.⁵⁰¹ I used a search for all cases tagged with a key number from the Contracts topic (95) as my control search in the denominator because I theorized that such a query would control for numerous differences between textualist and contextualist jurisdictions.⁵⁰² But this search could not control for any variations in contract law across states that impact the level of general contract litigation. Suppose there is more contract litigation in contextualist states because these jurisdictions have adopted most modern doctrines. This would lower the ratios (i.e., the percentages) established by my key number searches for those territories, distorting any comparison between the two sets of states. In other words, perhaps my finding that the ratio of interpretation cases to contracts cases does not vary by a statistically significant amount in textualist and contextualist territories simply reflects the fact that contextualist states have more of *both* types of litigation.

While this is an interesting explanation for my results, it is inconsistent with at least one aspect of the data: The totals for interpretation cases and for general contract cases were much higher in textualist states by every measure used. To illustrate, for query three in state and federal court, there were 2,548 cases in the textualist group and 847 in the contextualist group. For the control search in state and federal court, there were 5,436 cases in textualist states and 2,108 in contextualist states.⁵⁰³ These numbers do not preclude the possibility that differences in areas of contract law beyond interpretation distorted my results by increasing the number of cases in the denominator for the contextualist group. But they are a challenge that must be accounted for by such an explanation.

Fourth, beliefs about the law can influence behavior in ways that make it difficult to empirically measure the impact of variations in legal doctrine. Suppose, for example, that sophisticated contracting parties (and/or their attorneys) generally believe that contextualism raises enforcement costs in comparison to textualism. This might lead such parties to act in ways that increase the probability that textualist principles will govern their agreements. They could use tools like choice-of-law and choice-of-forum clauses, or they could locate more of their business activity in textualist

⁵⁰¹ See *supra* notes 445–63 and accompanying text.

⁵⁰² See *supra* notes 458–65 and accompanying text.

⁵⁰³ See *supra* Table 4.

states. Sophisticated parties would be particularly motivated to follow such practices for transactions that are more likely to result in an interpretive dispute. If this analysis is accurate, then a disproportionate amount of interpretation litigation will be commenced in textualist jurisdictions, inflating the ratio of interpretation cases to general contract cases in comparison to contextualist states.⁵⁰⁴ Alternatively, perhaps the view that contextualism increases uncertainty (and thus raises enforcement costs) induces sophisticated parties to settle disputes more quickly when they are embroiled in litigation in jurisdictions that follow the broader interpretive approach. This would lower the ratio of interpretation cases to contract cases in those territories vis-à-vis textualist states.⁵⁰⁵

Both of the theories discussed in the prior paragraph are intended to defend textualism against my findings. And my data provides some support for the first hypothesis. As discussed two paragraphs above, my Westlaw queries returned far more interpretation and general contract cases in textualist states than in contextualist states. That is consistent with businesses taking steps to have their contracts governed by textualist doctrine. But the analysis in each theory could easily be reversed to support contextualism.⁵⁰⁶ And other pro-contextualist explanations of my findings are imaginable. The key point is this: If contracting parties are changing their behavior because of their *beliefs* about the impacts of interpretation law, then it might be difficult or impossible to measure the *actual* impacts of the two interpretive approaches.⁵⁰⁷

⁵⁰⁴ My thanks to Omri Ben-Shahar of the University of Chicago School of Law for this suggestion.

⁵⁰⁵ My thanks to Russell Korobkin of the UCLA School of Law for this suggestion.

⁵⁰⁶ To illustrate, if parties believe that textualism increases enforcement costs, they might be more likely to settle lawsuits filed in textualist states, lowering the ratio of interpretation actions to contract actions in those jurisdictions.

⁵⁰⁷ Robert Thompson identified a similar problem in his empirical study of the factors that lead courts to pierce the corporate veil. He explained that any assessment of these factors “can be affected to the extent that litigants understand the prior learning on a legal issue and use that knowledge to decide which cases to file, to continue on appeal, or to settle.” Thompson, *supra* note 101, at 1046. He offered as an example undercapitalization. Courts frequently state that undercapitalization is the most important factor in deciding whether to pierce the veil. But if defendants are aware of this, then they are probably more likely to settle cases in which the corporate entity was undercapitalized. *Id.* at 1046 n.67. And this means that undercapitalization will not appear in the caselaw as often as other, less significant piercing factors. Therefore, an empirical finding that undercapitalization is seldom litigated in reported decisions

Given the methodological concerns presented in this subpart, textualists and contextualists are justified in harboring considerable doubts about my results. And there are likely multiple additional objections one could raise. But it is also possible that I failed to find a statistically significant difference in the level of enforcement costs between textualism and contextualism because there actually is no real difference between the two approaches when it comes to such costs. The next subpart addresses that possibility.

2. Alternative Explanations for the Study Findings

For thirteen of fourteen measures used in my study there was no statistically significant difference between textualism and contextualism. Perhaps that is because the number of lawsuits filed and the length of those proceedings are substantially the same under the two schools of contract interpretation. This subpart offers three potential explanations for why that might be the case.

Both textualists and contextualists have articulated multiple theories to justify the conclusion that their school of interpretation best reduces enforcement costs.⁵⁰⁸ One potential explanation for my findings is that all (or the bulk) of these theories are false. To illustrate, textualists assert that it is easier to establish the existence of an ambiguity under contextualism because parties have more material available out of which to craft reasonable understandings of the relevant contract language. That seems plausible enough. But maybe additional evidence typically does not seriously improve a claim that an agreement is ambiguous. The driving force in ambiguity determinations, even in contextualist states, might be the express terms of the contract. Similarly, contextualists maintain that textualism creates incentives to write longer contracts because parties are less able to rely upon extrinsic evidence should a dispute arise. And longer agreements are more likely to have contradicting terms, increasing litigation. This too is a plausible theory. But perhaps the incentives in textualist states are too weak to influence drafting practices. In particular,

(or in other cases) does not necessarily support the conclusion that undercapitalization is less important than as suggested by judicial pronouncements regarding piercing doctrine.

⁵⁰⁸ See *supra* Part IV.B.3.

parties might be more concerned about transaction costs than enforcement costs, and so they prefer to take their chances that litigation will result rather than spend time constructing longer agreements.⁵⁰⁹ Another possibility is that parties and their lawyers are quite proficient at drafting extensive contracts and so the predicted contradictions seldom materialize.

A second explanation for my findings is that there is considerable truth in all or most of the textualist and contextualist theories about enforcement costs, but the impacts of each approach largely cancel out. For example, textualists argue that contextualism promotes uncertainty because parties cannot know in advance which evidence a court is likely to find persuasive when deciding whether a contract is ambiguous. Indeed, until discovery commences, parties will often not even have access to all of the pertinent materials. This makes it difficult to predict the result of an ambiguity determination. Contextualists counter that textualism promotes uncertainty because the only inputs at the ambiguity stage under that approach are the contract and the judge, and judges vary dramatically in their acontextual understandings of contract language. Moreover, until a lawsuit is filed, the parties will not even know which judge is going to preside over their dispute. This also makes it difficult to predict the outcome of an ambiguity determination.

It is entirely conceivable that both of these theories are correct, but that the resulting levels of uncertainty are substantially equivalent. In other words, the uncertainty created by not knowing what evidence a contextualist judge will find persuasive could be substantially the same as the uncertainty created by not knowing how a textualist judge is going to view contract language in the absence of any extrinsic evidence—i.e., given only his or her background. One might respond that such equivalence is unlikely. But when all of the possible pathways to increased (or reduced) enforcement costs under each approach are added into the mix, it would not be surprising if the full panoply of countervailing forces sufficiently balance out such that there is no statistically significant difference between textualism and contextualism.

⁵⁰⁹ See *supra* notes 374–75 (discussing why parties might be more concerned with transaction costs than enforcement costs).

A third possibility is that the countless other factors that influence whether a lawsuit is filed and how long it lasts swamp any impact resulting from the interpretive approach in use by the courts. To illustrate, the inherent ambiguities in language might make litigation over the meaning of agreements extremely unpredictable regardless of which school of interpretation is employed. If that is the case, then the construction of contracts may be so difficult under either approach that changing between them has only a small effect on enforcement costs. Indeed, the general lack of predictability in interpretation cases is well known and has been cited as a basis for the belief that a shift towards contextualism will have no effect on the level of uncertainty parties face.⁵¹⁰ Furthermore, the problems inherent in language that infect interpretation litigation are only one of many factors that can influence whether a case is filed and the length of the proceeding. Others include (1) the rules of procedure and evidence, (2) the capacities of judges, lawyers, and jurors, (3) the nature of the parties (i.e., are they businesses or consumers), and (4) the relationship of the parties (i.e., are they long-term partners or was this their first deal). When all of these forces are considered, it makes sense to believe that even if the choice of interpretive approach matters to some degree, the impacts are too trivial to be measurable.

To recap, according to the first explanation, the underlying textualist and contextualist theories about enforcement costs are generally false. According to the second explanation, the underlying theories are largely true, but the impacts of each approach cancel out. And according to the third explanation, the theories are again largely true, but all of the other factors that influence enforcement cost levels swamp any difference between the two approaches. If one of these explanations is valid, then enforcement costs probably should no longer play a substantial role in debates over the best approach to contract construction.

⁵¹⁰ See McLaughlan, *supra* note 10, at 35 (“As I pointed out at the very beginning of this article, contract interpretation cases tend to be the most intractable of all contractual disputes and their outcome is notoriously difficult to predict. It is difficult to believe, therefore, that a more liberal approach to the reception of evidence of prior negotiations would result in any greater uncertainty. Indeed, in those cases where the evidence revealed that the parties formed a common intention as to the meaning of the words in dispute, the opposite might be the case.”).

IV. CONCLUSION

I had two general purposes in writing this article. First, I sought to explain how authors can employ the West Key Number System as a data collection and coding device to facilitate empirical scholarship, and why research conducted using key numbers is worthwhile. Second, I intended to contribute to the academic literature and caselaw regarding the optimal method of contract interpretation by executing the first empirical study analyzing the relationship of interpretive approach to enforcement costs.

Starting with the second purpose, the findings of my study suggest that enforcement costs do not vary between textualism and contextualism. This undercuts a signature claim advanced by textualist courts and commentators—that contextualism increases the level of interpretation litigation from the textualist baseline. And it similarly undermines the contextualist counterclaim that textualism is the approach with higher enforcement costs. Of course, this was a single study employing a methodology with considerable limitations. Further work is necessary before even tentative conclusions are possible about which approach to contract interpretation (if any) best minimizes the number of lawsuits filed and the length of those proceedings. But this paper contains the beginnings of an empirically-informed debate about contract interpretation enforcement costs.

I am currently investigating several potential follow-up projects that would further analyze the relationship of interpretive approach and enforcement costs. Each would employ key numbers for data collection and coding. First, I may conduct a more comprehensive study using essentially the same methodology employed in this article. The principal difference between the two projects is the quantity of key numbers involved. In the new work, I hope to include key numbers from all or most of the topics that concern the interpretation of agreements rather than just classifications contained in the Contracts and Evidence topics. However, the comprehensive study faces technological difficulties that might make its completion impracticable.⁵¹¹

⁵¹¹ For a summary of the problem, see *supra* note 449.

Second, I might rerun the queries from the current project in a larger sets of states. The principle roadblock to this approach is that it was quite difficult to identify ten states that I could fairly classify as either textualist or contextualist. Creating a larger data set will be even more challenging, if not impossible.

Third, I may attempt to run the same or modified searches in states before and after they shifted from one interpretive approach to another. In several jurisdictions there might be a seminal case in which the state supreme court changed the local approach to contract interpretation from textualism to contextualism.⁵¹² Comparing the ratio of interpretation litigation to general contract litigation before the shift to the ratio afterwards could provide helpful information regarding the impact of interpretive approach on enforcement costs. In particular, this protocol would eliminate the need to control for variations between states. The comparison would not be textualist states versus contextualist states during a single time period. Rather, it would be a set of states at time one (the textualism era) versus the same states at time two (the contextualist era).⁵¹³ This methodology would introduce new confounding variables—namely legal and social variations between the two time periods in each state. But combining the findings of such a study with those of my current project may provide a more complete picture of the effects textualism and contextualism have on litigation.

Fourth, Article 2 of the Uniform Commercial Code embraces contextualist interpretation.⁵¹⁴ This means that courts in textualist states are required to use contextualism when construing agreements for the sale of goods.⁵¹⁵ One possible study protocol focusing exclusively on textualist states would involve comparing the ratio of common law interpretation cases to common law contract cases with the ratio of U.C.C. interpretation cases to U.C.C. contract cases. Since the common law decisions in these states apply textualism and the U.C.C. decisions apply contextualism, this

⁵¹² See, e.g., *Pacific Gas & Electric Co. v. G.W. Thomas Drayage & Rigging Co.*, 442 P.2d 641 (Cal. 1968).

⁵¹³ This oversimplifies. There would be a time one and time two for each jurisdiction since states obviously changed from textualism to contextualism at different points.

⁵¹⁴ See U.C.C. §§ 1-303(d); *id.* 2-202 & cmt. 1(c).

⁵¹⁵ See *supra* note 453 for a discussion of a possible disjunction between what the U.C.C. requires and what courts in textualist states actually do.

type of comparison could highlight the impacts of each interpretive approach within a single state during a single time period. Thus, as with the third project, all data comparisons would be intrajurisdictional, nullifying the need to control for differences across state lines. And because the textualist and contextualist data for each state would come from the same era, legal and social changes would not be a problem with this protocol. But again there would be new confounding variables, including potential distinctions between common law litigation and statutory litigation and differences between the two areas of contract law (common law versus the U.C.C.). In addition, because the U.C.C. adopted other modern contract doctrines, using general contract litigation in the denominator would reintroduce an important problem from my current study.⁵¹⁶ Nonetheless, this project might contribute to obtaining a better overall understanding of the relationship between interpretive approach and enforcement costs.⁵¹⁷

As the reader can see, contract interpretation is a topic ripe for empirical analysis via the Key Number System. But the same is also true for legal issues arising in numerous other fields of law. Indeed, any time the quantity of litigation is relevant to a policy question or sociological inquiry, the data available through key number searches might provide helpful insights. To illustrate, litigation levels are frequently a factor in rule selection. One of the most common arguments in clashes about the appropriate legal norm is that a particular option will “open the litigation floodgates.”⁵¹⁸ An excellent general example is that increased litigation is allegedly one of the principal problems with adopting broad standards rather than specific rules as a governing norm.⁵¹⁹ Scholars can conduct empirical work employing key numbers to assess such assertions about

⁵¹⁶ See *supra* notes 498–503 and accompanying text.

⁵¹⁷ Note also that one could conduct a study mixing key numbers with other information. For example, one could analyze the ratio of interpretation cases tagged with a pertinent key number to some non-key number denominator, such as civil filings denoted as contracts cases by administrative offices of the courts, assuming such data exists.

⁵¹⁸ Searches on Westlaw run on September 2, 2015, returned 1,460 journal articles and 557 state and federal cases containing “floodgates” within four words of “litigation.” See also Ellie Margolis, *Closing the Floodgates: Making Persuasive Policy Arguments in Appellate Briefs*, 62 MONT. L. REV. 59, 73 n.74 (2001) (explaining that the litigation floodgates argument is “much overused”); Michael R. Smith, *The Sociological and Cognitive Dimensions of Policy-Based Persuasion*, 22 J.L. & POL’Y 35, 58 (2013) (noting that the litigation floodgates argument is “popular”).

⁵¹⁹ See *supra* note 500 and accompanying text.

litigation quantities in countless circumstances, as I have here with the textualist and contextualist theses concerning enforcement costs.

The Key Number System also makes it easier to answer calls for certain types of empirical research long neglected by law professors. For example, Hall and Wright note that legal scholars have “largely overlooked” studies that treat “case law as an *independent* variable, meaning they ask how case law influences other social and economic conditions.”⁵²⁰ This is surprising, they explain, because, “[w]ith its diverse ‘laboratory of states,’ the United States offers boundless opportunities to learn from the ‘natural experiments’ created by the inevitable differences in case law among jurisdictions and over time.”⁵²¹ Korobkin highlights this problem in the contracts context, observing that scholars seldom investigate empirically whether rival rules of contract law have different impacts on the behavior of contracting parties.⁵²² He argues that the “measurable differences in doctrine between jurisdictions” create opportunities for comparing the impacts of competing approaches.⁵²³ That, of course, is precisely what I attempted to do via my study of the effects of textualist and contextualist interpretation on enforcement costs.⁵²⁴ And the fourth potential follow-up project discussed above—contrasting litigation levels under the common law with those under the U.C.C.—is essentially concerned with the same type of comparison. Korobkin further explains that “changes in doctrine over time can make it possible to study the effects of doctrine by comparing behavior in the same jurisdiction before and after a landmark case, series of cases, or statutory change.”⁵²⁵ That is what I am

⁵²⁰ Hall & Wright, *supra* note 3, at 86.

⁵²¹ *Id.* at 87.

⁵²² Korobkin, *supra* note 4, at 1062; *see also* Crystal, *supra* note 56, at 306 (“Other internal empirical studies could focus on the operation of rules. By gathering data about the relative percentage of types of litigated cases, it would be possible to identify those rules of law that are frequently litigated. Focus could then turn to normative questions. Is the amount of litigation of a particular rule or doctrine a matter of concern? Are there alternative dispute resolution mechanisms that could be developed?”).

⁵²³ Korobkin, *supra* note 4, at 1063.

⁵²⁴ *See also* Richard Danzig, *Comments on Professor White’s Paper*, 1988 ANN. SURV. AM. L. 49, 59 (1988) (“One then asks about the anvil, how does it affect the disposition of real cases in an everyday way? How does it affect the settlement process, etc.?”).

⁵²⁵ Korobkin, *supra* note 4, at 1063; *accord* Geis, *supra* note 4, at 492 (“The basic idea here is to find a situation where there has been a doctrinal change—perhaps due to a landmark court decision or statutory enactment—and measure the impact of the change on subsequent contract terms in the jurisdiction. Parallel jurisdictions might also be included as a form of control. This would be analogous

proposing to do in the third follow-up project—contrasting enforcement costs during an earlier textualist period with those during a later contextualist period.

Remember also that key number projects can be completed expeditiously and with minimal resources. Indeed, they require substantially less time and money than most other types of empirical work. Scholars thus can execute the types of projects called for by Hall, Wright, Korobkin, and others with relative ease.

Of course, the methodological problems inherent in using the American Digest System for data collection and coding undercut the value of this type of research. But empirical work need not be perfect to possess significant value.⁵²⁶ This is especially true in fields such as law and public policy where decision-makers often must act despite “profound empirical uncertainty.”⁵²⁷ Practitioners in the social and hard sciences normally have the option of awaiting better evidence before reaching a conclusion. But deferring on legal and policy questions is frequently impossible. Judges must adjudicate the cases that come before them.⁵²⁸ And when legislators postpone a decision, they simply leave the status quo legal rule in place.⁵²⁹ When choice is unavoidable, greater methodological rigor can be an unaffordable luxury. Accordingly, while the research protocol recommended in this article suffers from many limitations, the findings of key number studies are worthy of consideration by courts, legislatures, and scholars, especially when such studies address issues for which there is no better evidence.

West’s American Digest System is a treasure trove of coded data. My hope is that this article inspires scholars in every field of law to mine that data in order to analyze countless empirical questions of fundamental importance to our nation’s legal system.

to an event study in finance or economics and might help to shed light on the incentives and effects of differing legal default rules.”).

⁵²⁶ See *supra* notes 209–34 and accompanying text.

⁵²⁷ Goldsmith & Vermeule, *supra* note 230, at 154.

⁵²⁸ Thomas C. Grey, *Langdell’s Orthodoxy*, 45 U. PITT. L. REV. 1, 25 (1983).

⁵²⁹ For a contracts example illustrating this principle, see Geis, *supra* note 4, at 492 (“After all, lawmakers cannot abandon the job; there must be some legal rule or standard that takes effect in the absence of stated contractual intentions. Even a decision to eschew default rules or, analogously, to strike contracts void for indefiniteness whenever undocumented contingencies arise is itself a default rule.”).

APPENDIX 1: WESTLAW SEARCHES USED IN THE STUDY

This appendix sets forth the searches I ran on Westlaw to collect the data for my study. Recall that I organized the key numbers chosen for the study into three groups: (1) contract interpretation key numbers from the Contracts digest topic (topic 95), (2) contract interpretation key numbers from the Evidence digest topic (topic 157), and (3) parol evidence rule key numbers from the Evidence topic. The search terms for each group are listed below.

Next, I constructed three search queries for each state. The first query contained only the key numbers in group (1). The second contained the key numbers in groups (1) and (2). And the third contained the key numbers in all three groups—(1), (2), and (3). The three queries were run in both the relevant state database and the relevant mixed state/federal database. Thus, there were six searches in total for each jurisdiction. The results of these searches served as proxies for the level of contract *interpretation* litigation in each state during the time period of the study.

The three queries were run in exactly the same form in every state database and each of the three is presented below. Running the queries in the mixed state/federal databases required additional search language to limit the cases retrieved to the appropriate jurisdiction. Accordingly, I have set forth the three queries as run in the Indiana state/federal databases to highlight the limiting language that was necessary.

I also ran a control search for all cases classified with a contracts key number. This search served as a proxy for the total level of *contract* litigation in each state during the time period of the study. Like the three interpretation queries, the control required additional restrictive language when it was run in the mixed state/federal databases. I thus have set forth both versions of the control search—state *and* state/federal—again using Indiana as an example of the latter.

Finally, all queries were run with a date restriction that limited the dataset to cases decided from January 1, 2000 through December 31, 2009. The search language restricting the results to this time period is contained in the queries below.

Group 1: Topic 95 (Contracts), Section II, Subsection (A)—Selected Key Numbers.

95k143 95k143.5 95k147 95k148 95k149 95k150 95k151 95k160 95k161
95k162 95k163 95k164 95k165 95k166 95k167 95k169 95k170 95k171
95k172 95k173 95k174 95k175 95k176

Group 2: Topic 157 (Evidence), Section XI, Subsection (D)—All Key Numbers.

157XI(D)⁵³⁰

Group 3: Topic 157 (Evidence), Section XI, Subsection (A)—All Key Numbers *and* Subsection (C)—Selected Key Numbers.

157XI(A)

157k439 157k440 157k441 157k442 157k443 157k444⁵³¹

Query 1—State (Group 1).

(95k143 95k143.5 95k147 95k148 95k149 95k150 95k151 95k160 95k161
95k162 95k163 95k164 95k165 95k166 95k167 95k169 95k170 95k171
95k172 95k173 95k174 95k175 95k176) & da(aft 12/31/1999) & da(bef
01/01/2010)

Query 2—State (Groups 1 & 2).

(157XI(D) (95k143 95k143.5 95k147 95k148 95k149 95k150 95k151
95k160 95k161 95k162 95k163 95k164 95k165 95k166 95k167 95k169
95k170 95k171 95k172 95k173 95k174 95k175 95k176)) & da(aft
12/31/1999) & da(bef 01/01/2010)

Query 3—State (Groups 1, 2 & 3).

(157XI(D) (95k143 95k143.5 95k147 95k148 95k149 95k150 95k151
95k160 95k161 95k162 95k163 95k164 95k165 95k166 95k167 95k169
95k170 95k171 95k172 95k173 95k174 95k175 95k176) (157XI(A)
157k439 157k440 157k441 157k442 157k443 157k444)) & da(aft
12/31/1999) & da(bef 01/01/2010)⁵³²

⁵³⁰ Using this term in a search will retrieve all cases with any key number contained in this section of the Evidence topic outline.

⁵³¹ These are the selected key numbers from topic 157, section XI, subsection (C).

⁵³² Note that some of the parentheses in these searches are superfluous. I included them when I ran the searches because they make it easier to see precisely what I was searching for without altering the results retrieved.

Query 1—State & Federal (Group 1).

Here is the search I ran for Indiana state and federal cases in the database IN-CS-ALL:

co(in) & (95k143 95k143.5 95k147 95k148 95k149 95k150 95k151 95k160 95k161 95k162 95k163 95k164 95k165 95k166 95k167 95k169 95k170 95k171 95k172 95k173 95k174 95k175 95k176) & da(aft 12/31/1999) & da(bef 01/01/2010)

Notice the italicized language at the start of the search, which was not included in query 1 when I searched only for state cases. The Westlaw database IN-CS-ALL contained “[c]ases from Indiana state courts, the U.S. Supreme Court, federal appellate courts authoritative in the Seventh Circuit, U.S. bankruptcy and district courts in Indiana, and the Judicial Panel on Multidistrict Litigation.”⁵³³ A search in that database using query 1 without the italicized language would have retrieved numerous cases that did not originate in Indiana, such as Seventh Circuit cases that were initially filed in Illinois or Wisconsin. The italicized language (substantially) restricted the search to cases arising in Indiana. The addition of that language is significant. Compare the following results:

Query 1—State (run in IN-CS): 229 cases

Query 1—State & Federal (run in IN-CS-ALL *with* the limiting language): 280 cases

Query 1—State & Federal (run in IN-CS-ALL *without* the limiting language): 372 cases

Without the limiting language, the state/federal search would have retrieved an additional 92 cases for Indiana that were irrelevant to my research hypotheses. Comparable language was used in all state/federal searches. For example, I included “co(ny)” for the New York queries and “co(ca)” for the California queries.⁵³⁴

⁵³³ This was the Westlaw.com description of the IN-CS-ALL database. Note that accessing databases works differently at WestlawNext.com. To search all Indiana cases, one would now access the dropdown menu next to the search box and select “Indiana” and “Include Related Federal,” and confirm that nothing else is selected.

⁵³⁴ Note that WestlawNext now uses the letters “ca” to refer to California *and* to courts of appeal. Thus, to restrict a search to federal cases arising in California, the appropriate court restrictor is co(#ca). The pound sign ensures that WestlawNext reads “ca” as referring to a state rather than a type of court.

Query 2—State & Federal (Group 1 & 2).

co(in) & (157XI(D) (95k143 95k143.5 95k147 95k148 95k149 95k150 95k151 95k160 95k161 95k162 95k163 95k164 95k165 95k166 95k167 95k169 95k170 95k171 95k172 95k173 95k174 95k175 95k176)) & da(aft 12/31/1999) & da(bef 01/01/2010)

Query 3—State & Federal (Group 1, 2, & 3).

co(in) & (157XI(D) (95k143 95k143.5 95k147 95k148 95k149 95k150 95k151 95k160 95k161 95k162 95k163 95k164 95k165 95k166 95k167 95k169 95k170 95k171 95k172 95k173 95k174 95k175 95k176) (157XI(A) 157k439 157k440 157k441 157k442 157k443 157k444)) & da(aft 12/31/1999) & da(bef 01/01/2010)

Control Search—State.

to(95) & da(aft 12/31/1999) & da(bef 01/01/2010)⁵³⁵

Control Search—State & Federal.

co(in) & to(95) & da(aft 12/31/1999) & da(bef 01/01/2010)

⁵³⁵ This search retrieves all cases tagged with any topic 95 (Contracts) key number in the relevant database.

APPENDIX 2: STATISTICAL ANALYSIS

A. t-tests

My study compared the ratio of contract interpretation litigation to general contract litigation in textualist states to the same ratio in contextualist states. The ratio of interpretation litigation to general contract litigation was calculated in twelve distinct ways for each of the ten jurisdictions in the study. The statisticians from the University of Georgia Statistical Consulting Center ran t-tests on each of these twelve measures. The t-tests found no statistically significant difference for eleven of the twelve. On the twelfth measure—query 3, unweighted, state and federal—the higher level of litigation in textualist states was statistically significant at the .05 level.

Table 6 presents the t-test results for each of the twelve comparisons. The first three columns in the table list the means, the ninety-five percent confidence intervals, and the standard deviations for the two groups of states. Those columns also include the differences between the two sets of states on each of these benchmarks. The last two columns contain the t-value and the p-value. Note that the method (pooled), the variances (equal), and the degrees of freedom (eight) are the same for each of the twelve t-tests. Thus, those parameters were not included in the table.

The t-tests concern two relevant populations—all textualist states and all contextualist states. The five states in my textualist group and the five in my contextualist group are samples taken from these populations. However, a t-test presumes that the samples were chosen randomly from the population. That is not the case here. I exercised my judgment in selecting the states for my two sample sets.⁵³⁶ This limits the validity of the t-tests.

Note that the lack of random sampling in selecting states for the study is distinct from the problem of non-random sampling discussed earlier in this paper. Previously, I explained that virtually all cases that receive key number treatment are reported and that reported decisions are probably not a random subset of all litigated matters.⁵³⁷ This means that there is a high

⁵³⁶ See *supra* notes 443–44 and accompanying text.

⁵³⁷ See *supra* notes 148–49, 168–205 and accompanying text.

likelihood that the t-tests are infected with two types of selection bias: (1) the mostly reported cases used to calculate the relevant ratios in each state are not representative of all cases in those states; and (2) the jurisdictions chosen for the study are not representative of all states that employ textualist and contextualist interpretation.

The p-value of the t-test is the probability that the data would produce differences as large as those seen between the samples here (between the two sets of five states) if it is assumed that there is no true difference between the textualist and contextualist populations (all textualist states and all contextualist states). When the p-value is smaller than a preset significance level, sufficient evidence exists to state that there are statistically significant differences between the textualist and contextualist approaches. If the p-value is larger than the preset significance level, then the two approaches are not significantly different. The typical significance level is .05 and that was the level used in the t-tests for this study. When the t-test reaches a result at or below the .05 level, it can be stated with ninety-five percent confidence that the differences in the two samples reflect genuine differences in the populations.⁵³⁸

To explain further, the t-tests start from the assumption that the two populations—all textualist states and all contextualist states—are not different, and thus have comparable means. In other words, the average ratio of interpretation litigation to contract litigation in textualist states is the same as the average ratio in contextualist states. If the population means are the same, we would expect the sample means (the average ratios from my two groups of five states) to be largely the same.⁵³⁹ Given that assumption, the differences between the means of the two samples must be explained. Since the sample means are different, it is possible that the

⁵³⁸ See LAWLESS ET AL., *supra* note 19, at 233 (“Traditionally, when a result has a 5 percent or less chance of occurring but occurs nonetheless, researchers consider the result to be statistically important. The reason for choosing 5 percent is largely because in a normal distribution . . . approximately 95 percent of the observations fall within two standard deviations above and below the mean. Thus, an observation more than two standard deviations above or below the mean has only a 5 percent chance of being observed. This has become a norm for what counts as a meaningful statistical outcome.”).

⁵³⁹ When samples are taken from two populations with identical means, the expectation is that the means of the samples will be similar, not identical, because there is variability among the members of each population—i.e., because the numerical values of the population members are spread out to some degree.

populations are different—i.e., that some difference in the populations is causing the ratio of interpretation cases to contract cases to be higher in textualist states. However, the t-test results indicate that the sample means are *not* sufficiently different to rise to the level of statistical significance for eleven of the twelve measures used in the study. Thus, we cannot say, with appropriate confidence, that the means of the two populations are actually different by those eleven measures. In sum, the conclusion of the t-tests is that even if the population means of textualist and contextualist states are identical, it would not be surprising to find that the means of the two samples differ by the amount reflected in my data (again, for eleven of twelve measures).

Table 6. t-Tests

Legend: C = Contextualist; T = Textualist; Q1 = Query 1; Q2 = Query 2; Q3 = Query 3; S = State; S/F = State & Federal; W = Weighted; U = Unweighted. Bold Italics = statistically significant result.

Measure	Mean	95% CL Lower	Mean Upper	Std. Dev.	t value	p value
C-Q1-S-W	0.3549	0.2694	0.4404	1.1087		
T-Q1-S-W	0.4203	0.3096	0.5310	2.3511		
Difference	-0.0654	-0.2033	0.0725	1.8381	-1.09	0.306
C-Q2-S-W	0.4159	0.3503	0.4815	0.8507		
T-Q2-S-W	0.4618	0.3653	0.5583	2.0496		
Difference	-0.0459	-0.1636	0.0718	1.5691	-0.90	0.3948
C-Q3-S-W	0.4410	0.3864	0.4955	0.7078		
T-Q3-S-W	0.4908	0.4037	0.5780	1.8505		
Difference	-0.0499	-0.1550	0.0552	1.401	-1.09	0.3056
C-Q1-S-U	0.3738	0.2850	0.4626	0.0715		
T-Q1-S-U	0.4478	0.3467	0.5489	0.0814		
Difference	-0.074	-0.1858	0.0378	0.0766	-1.53	0.1653
C-Q2-S-U	0.4240	0.3486	0.4994	0.0607		
T-Q2-S-U	0.4868	0.3984	0.5752	0.0712		
Difference	-0.0628	-0.1593	0.0337	0.0662	-1.50	0.1719
C-Q3-S-U	0.4408	0.3698	0.5118	0.0572		
T-Q3-S-U	0.5100	0.4288	0.5912	0.0654		
Difference	-0.0692	-0.1588	0.0204	0.0615	-1.78	0.1129
C-Q1-S/F-W	0.3168	0.2437	0.3899	1.2091		
T-Q1-S/F-W	0.3969	0.3014	0.4923	2.5343		
Difference	-0.08	-0.1975	0.0374	1.9855	-1.57	0.1548

Measure	Mean	95% CL Mean		Std. Dev.	t value	p value
		Lower	Upper			
C-Q2-S/F-W	0.3743	0.3071	0.4416	1.1118		
T-Q2-S/F-W	0.4383	0.3531	0.5236	2.2650		
Difference	-0.0640	-0.1696	0.0416	1.7842	-1.40	0.1996
C-Q3-S/F-W	0.4018	0.3408	0.4627	1.0078		
T-Q3-S/F-W	0.4688	0.3912	0.5464	2.0611		
Difference	-0.0670	-0.1630	0.0290	1.6223	-1.61	0.1461
C-Q1-S/F-U	0.3338	0.2524	0.4152	0.0655		
T-Q1-S/F-U	0.4260	0.3410	0.5110	0.0685		
Difference	-0.0922	-0.19	0.00555	0.067	-2.18	0.0613
C-Q2-S/F-U	0.3816	0.2990	0.4642	0.0665		
T-Q2-S/F-U	0.4664	0.3911	0.5417	0.0606		
Difference	-0.0848	-0.1776	0.00801	0.0636	-2.11	0.0682
C-Q3-S/F-U	0.4006	0.3206	0.4806	0.0644		
T-Q3-S/F-U	0.4950	0.4277	0.5623	0.0542		
Difference	-0.0944	-0.1812	-0.00757	0.0595	-2.51	0.0365

B. Logistic Regressions

The Georgia statisticians also ran mixed-effect logistic regression models on two of my comparisons between textualist and contextualist states—(1) query 1, unweighted, state courts, and (2) query 1, unweighted, state and federal courts. The two logistic regressions found no statistically significant difference between the contextualist and textualist jurisdictions.

Let me begin by briefly explaining why the logistic regressions were run for only two of the twelve comparisons. First, any comparisons involving queries 2 and 3 could not be included because the type of regression used assumes that the ratio of contract interpretation cases to contract cases can never be greater than one (or negative). That is true when using query 1 because query 1 contained *selected* key numbers from topic 95 and the control search used *all* of the key numbers from topic 95. But

queries 2 and 3 included numbers from both topic 95 *and* topic 157. Thus, in theory, the raw totals for queries 2 and 3 could be larger than the control search in a given circumstance. While that is highly unlikely in practice, the fact that it is possible made it improper to run mixed-effect logistic regression models for the eight comparisons using queries 2 and 3. Second, the goal of the study was to analyze the variability across states with regard to the ratio of contract interpretation litigation to general contract litigation. The comparison of textualism to contextualism was done at the state level. The number of observations (i.e., cases) within each jurisdiction is ignored in such a comparison; the observations are only used to calculate the ratios/percentages for each territory. This means that each state counted equally towards the comparison in the mixed-effect logistic regression models. And that disqualified the weighted query 1 comparisons since, in those comparisons, the states do not count equally. Instead, states with a greater number of observations count for more.

Next, note that once again the selection bias created by not using random sampling limits the validity of the logistic regressions. In other words, the results of the two regressions do not necessarily provide a valid answer to the question about the effects of textualism and contextualism because of the sampling techniques employed.

The statistical model for this study is modelling the probability of a case being contract interpretation litigation given that it is contract litigation, based on the data from the ten states in the two interpretation groups. The predictor included in this model is the method of interpretation—textualist or contextualist—with textualist serving as the baseline.

Logistic Regression 1—Query 1, Weighted, State.

Table 7. Fit Statistics

-2 Res Log Pseudo-Likelihood	8.15
Generalized Chi-Square	7.64
Gener. Chi-/Square/DF	<i>0.95</i>

Typically, if the italicized value in Table 7 is close to 1, this indicates that the model is an acceptable fit for the data.

Table 8. Covariance Parameter Estimates

Cov. Parm.	Subject	Estimate	Standard Error
Intercept	State	<i>0.09685</i>	0.05387

The italicized value in Table 8 is an estimate of the variability among the different states.

Table 9. Solution for Fixed Effects

Effect	Interpret	Estimate	Std. Error	DF	t value	P-value
Intercept		-0.2229	0.1464	7.776	-1.52	0.1674
Interpret	Contextualist	-0.3035	0.2131	8	-1.42	0.1921
Interpret	Textualist	0				

The “Estimate” column in Table 9 gives the estimated effects of the method of interpretation. In particular, the -0.3035 estimate means that the log odds of p —the probability of a case being a contract interpretation case given that it is a contract litigation case—decreases by -0.3035 on average when changing the interpretive method from textualist to contextualist. The log odds is a function of p , that is $\log(p/1-p)$. In sum, using contextualist interpretation will make p decrease. But according to the P-value (0.1921), this decrease is not statistically significant.

Table 10. Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	P-Value
Interpret	1	8	2.03	0.1921

The output from Tables 7 through 10 identifies the effect of the method of interpretation (contextualist or textualist) on the probability of a case being contract interpretation litigation given that it is contract litigation. Ignoring all problems with the data and assuming the pertinent assumptions are met, the result in these tables shows that there is no significant relationship between the method of interpretation and the probability of a case being contract interpretation litigation given that it is contract litigation. In other words, a state being contextualist or textualist does not have a significant association with a higher or lower proportion of contract interpretation litigation among all instances of contract litigation.

Logistic Regression 2—Query 1, Weighted, State & Federal.

The same basic explanations of the tables in the first regression also apply to the second regression. Thus, there is no elaboration on the meaning of the results in Tables 11 through 14.

Table 11. Fit Statistics

-2 Res Log Pseudo-Likelihood	6.13
Generalized Chi-Square	7.83
Gener. Chi-/Square/DF	0.98

Table 12. Covariance Parameter Estimates

Cov. Parm.	Subject	Estimate	Standard Error
Intercept	State	0.075222	0.04181

Table 13. Solution for Fixed Effects

Effect	Interpret	Estimate	Std. Error	DF	t value	P-value
Intercept		-0.3065	0.1278	7.559	-2.4	0.045
Interpret	Contextualist	-0.3955	0.1855	8	-2.13	0.0655
Interpret	Textualist	0				

Table 14. Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	P-Value
Interpret	1	8	4.55	0.0655