March 2016

Bargaining in the Shadow of Big Data

Dru Stevenson

Nicholas J. Wagoner

Florida Law Review

Founded 1948

BARGAINING IN THE SHADOW OF BIG DATA

Dru Stevenson* & Nicholas J. Wagoner**

Abstract

Attorney bargaining has traditionally taken place in the shadow of trial as litigants adjust tactics—and their inclination to negotiate a settlement—based on their forecast of the outcome of a trial and its associated costs. Lawyers bargaining on the verge of trial have traditionally relied on their intuition, knowledge of precedent, and previous interactions with the presiding judge and opposing counsel to forecast trial outcomes and litigation costs. Today, however, technology that leverages legal data is moving the practice of law into the shadow of the trends and patterns apparent in aggregated litigation data. This Article describes the tools that facilitate this paradigm shift and examines how lawyers use these tools to forecast litigation outcomes and reduce Coasean bargaining costs in both litigation and transactional fields. This Article also explores some of the risks associated with bargaining in the shadow of big data, and it offers guidance to lawyers leveraging these tools to improve their practice.

This discussion pushes beyond the cartoonish image of big data as a mechanical fortuneteller—predicting who will win or lose a case, supposedly eliminating research or deliberation. This Article also debunks the alarmist clichés about newfangled technologies eliminating jobs. Demand for lawyers who are capable of effective bargaining when confronted by big data will continue to increase as the legal profession catches up to the data-centric approach found in other industries. Ultimately, this Article paints a portrait of what big data really means for practicing attorneys, and it provides a framework for exploring the theoretical implications of lawyering in the era of information analytics.

INTRODUCTION

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

—Justice Oliver Wendell Holmes Jr.[1]

The practice of law includes prediction. Clients expect their lawyers to answer questions like "What are the odds of winning this case, and what is it going to cost me?"[2] Justice Oliver Wendell Holmes Jr. observed this fact more than a century ago when he wrote that "[t]he prophecies of what the courts will do in fact, and nothing more pretentious, are what I mean by the law."[3]

Today, practitioners still resemble Holmes's "black-letter men" of yesteryear more than they resemble the number-crunching masters of economics that Holmes had envisioned. The vast majority of attorneys still rely solely on their own experience, knowledge of case precedents, and intuition to predict what the courts will actually do.[4] But as mere mortals, attorneys—even the exceptional ones—are inherently limited in their capacity to retain and process the information necessary to make well-informed judgments. Likewise, attorneys have a limited range of personal experience. Computers, on the other hand, while lacking the ability to frame interesting questions or draw conclusions, are far better at storing, processing, and summarizing large volumes of information. Thus, by leveraging the quantitative strength of computers, lawyers can more accurately forecast how events will play out in litigation. This foresight, in turn, will allow lawyers and their clients to avoid costly mistakes, better appreciate the strengths and weaknesses of their cases, and increase their odds of securing favorable outcomes.

Others in the legal academy have already begun discussing the value of basing decisions on big data, though not necessarily in the context of lawyering. Professor Ian Ayres recently popularized the term "super crunching," referring to the statistical analysis of big data to guide real-

---

will facilitate lawyering in the shadow of data. I would also like to thank Keith Lee and Josh Blackman for providing thoughtful commentary on earlier drafts of this Article.

1. O.W. Holmes, *The Path of the Law*, 10 HARV. L. REV. 457, 469 (1897).

2. Tam Harbert, *Big Data Meets Big Law*, L. TECH. NEWS (Dec. 27, 2012), http://www.legaltechnews.com/id=1202555605051/Big-Data-Meets-Big-Law?slreturn=201502 23103023 (LEXIS subscription required); *see also* Daniel Martin Katz, *Quantitative Legal Prediction—or—How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry*, 62 EMORY L. J. 909, 912 (2013) ("Do I have a case? What is our likely exposure? How much is this going to cost? What will happen if we leave this particular provision out of this contract? How can we best staff this particular legal matter? These are core questions asked by sophisticated clients such as general counsels, as well as consumers at the retail level."); Farhad Manjoo, *Will Robots Steal Your Job?: Software Could Kill Lawyers*, SLATE (Sept. 29, 2011, 2:42 AM), http://www.slate.com/articles/technology/robot_invasion/2011/09/will _robots_steal_your_job_5.html ("The legal industry is one of the few remaining outposts of the corporate world whose operations are dictated mainly by human experience. Basic questions that anyone would want to know before committing to a million-dollar case—*How likely is it that I'll win? How good are my lawyers? Should I settle?*—can't be answered with certainty.").

3. Holmes, *supra* note 1, at 461.

4. *See* CLARENCE MORRIS, HOW LAWYERS THINK 11–19 (1938).

world decisions.[5] Super crunching improves decision-making by leveraging the quantitative strength of computers so that human users may spend more time doing what they do best—identifying relevant data to crunch, formulating questions to ask of their data, and then acting on the results.[6]

In popular culture and semiotics, *Moneyball*[7] has become the vernacular shorthand for data-driven decisions. In 2002, after making a fortune in finance by analyzing data to uncover and exploit inefficiencies in the financial markets, John Henry led a group of investors in the purchase of the Boston Red Sox baseball franchise.[8] Using the data-driven principles that Professor Ayres would later espouse in *Super Crunchers*, Henry built a team that went on to win the 2004 World Series and break the "Curse of the Bambino."[9] In a letter to a friend, he explained: "People in both fields operate with beliefs and biases. To the extent you can eliminate both and replace them with data, you gain a clear advantage."[10]

Henry's insight applies equally to the practice of law.[11] Lawyers, just like decision makers in every other field, operate with beliefs and biases that can cloud their judgment.[12] Yet more than a century ago, Justice Holmes pointed out a better path to predicting the law than that traveled by today's "black-letter man," a path that leads to a future where "the rational study of the law" requires a "man of statistics and master of economics."[13] Like Henry's Boston Red Sox, lawyers who embrace data-driven decision-making will gain a clear advantage over their

---

5. IAN AYRES, SUPER CRUNCHERS 10 (2007).

6. *See id.*

7. MICHAEL M. LEWIS, MONEYBALL (2003).

8. DAN SHAUGHNESSY, REVERSING THE CURSE: INSIDE THE 2004 BOSTON RED SOX 60 (2005).

9. *Id.* at 1.

10. LEWIS, *supra* note 7, at 90–91.

11. *See generally* Charles J. Snyder, *Moneyball Lawyering*, 65 ARK. L. REV. 837, 840–41 (2012).

12. *See* Anthony V. Alfieri, *(Un)Covering Identity in Civil Rights and Poverty Law*, 121 HARV. L. REV. 805, 826 (2008); Albert W. Alschuler, *The Defense Attorney's Role in Plea Bargaining*, 84 YALE L.J. 1179, 1221 (1975) (discussing how defense attorneys' biases and rapport with prosecutors can influence outcomes for defendants); Stephanos Bibas, *Plea Bargaining Outside the Shadow of Trial*, 117 HARV. L. REV. 2463, 2528–29 (2004); Darryl K. Brown, *Rationing Criminal Defense Entitlements: An Argument from Institutional Design*, 104 COLUM. L. REV. 801, 827 (2004); Marjorie Anne McDiarmid, *Lawyer Decision Making: The Problem of Prediction*, 1992 WIS. L. REV. 1847, 1868–75 (demonstrating that lawyers' decision-making is affected by cognitive bias); Jeffrey J. Rachlinski, *The Uncertain Psychological Case for Paternalism*, 97 NW. U. L. REV. 1165, 1217 n.274 (2003) (discussing cognitive biases for lawyers); Adrian Vermeule, *Should We Have Lay Justices?*, 59 STAN. L. REV. 1569, 1588–89 (2007); Adam S. Zimmerman, *Funding Irrationality*, 59 DUKE L.J. 1105, 1172 n.288 (2010). *But see* Oren Gazal-Ayal & Avishalom Tor, *The Innocence Effect*, 62 DUKE L.J. 339, 345–46 (2012).

13. Holmes, *supra* note 1, at 469.

counterparts who still cling to their Dictaphones, feather-quilled pens, and thumb-in-the-air predictions when handling their clients' legal affairs.

Why then, has most of the legal profession not yet found the path envisioned by Justice Holmes? The profession lags behind most other industries in crunching data to improve its marketing and delivery of services. The historical inaccessibility of litigation data and the inadequacy of the tools with which lawyers presently conduct legal research have created a significant bottleneck to progress.[14]

This is not to say that legal practitioners have not embraced modern technology. To the contrary, law firms were early adopters of office computers and word processors, scheduling and billing software, in-house networks, conference phones, and online databases like Lexis and Westlaw. Before smartphones became ubiquitous, lawyers commonly had PDA's to keep track of appointments, hourly billing, and client contacts.

But "big data" is different. Previous technological advances mostly made firms more efficient at the tasks that they were already doing—scheduling meetings, drafting documents, sharing ideas, and looking up cases. A few technologies considerably changed how lawyers approached a task. The most obvious example of this was the profession's massive shift in the early 1980s toward precision-timed billing—in minutes or fractions of hours—rather than ball-parked or "scheduled" fees once computer software made such time-tracking more feasible.[15] For the most part, however, the underlying nature of the work remained largely the same—researching, writing, and meeting with clients and opposing counsel—and the technology merely made these tasks more convenient or allowed lawyers to handle more cases. Even the research that attorneys now perform through Lexis and Westlaw is analytically analogous to the old approach in law libraries—finding cases in bound digests and reports as compared to using intricate indexing systems like West's Keys or

---

14. *See* Anastasia Boyko, *The Costs of (Inefficient) Legal Services Delivery*, N.Y. Sт. B. Ass'n J., Sept. 2012, at 34, 36.

15. *See* Mortimer D. Schwartz et al., Problems in Legal Ethics 127 (6th ed. 2004); Ed Wesemann, *Full Circle: Alternative Pricing*, 70 Or. Sт. B. Bull. 38, 38 (2010) ("The popularity of billing by the hour was supported by technology, first through 'one-write' record books, then electronic accounting machines, and eventually, computers until it became the pricing standard for the legal profession."); H. Edward Wesemann, *The Power of Price*, 23 Legal Mgmt. 22, 34–35 (2004) ("The billing-by-the-hour trend didn't earnestly begin until the 1960s, and it took computers to bring firms to complete dependence on hourly billing."). Ironically, even as computers facilitated the entrenchment of hourly billing, the advent of computers prompted leaders in the legal profession to predict that computers would also force the demise of time-based fees, as the new efficiency would make legal tasks too brief to generate sufficient time-based revenue. *See, e.g.*, Sandra L. Yost, *Alternative Billing Strategies—New Wave or Passing Fad?*, 13 Legal Mgmt. 18, 20 (1994) (discussing comments by state bar leaders at panel discussions about the impact of computers on alternate billing strategies).

Lexis's Headnotes.

Big data, by contrast, invites lawyers to make a fundamental change in their approach to the law itself by looking to statistical patterns, predictors, and correlations, in addition to the legal rules that purportedly control outcomes—case law, statutory law, procedural rules, and administrative regulations. Traditional lawyering required knowledge of the pertinent legal rules and the ability to apply them to a given set of facts, whether in litigation or in transactional work. This application of law to facts would yield an estimate about probabilities: that is, a prediction of the likelihood that a given rule would govern a given scenario. The question was whether a feature of the client's current situation would trigger a rule and its mandatory result. Analogies, comparisons, and normative judgments all figured into this assessment. A lawyer's fees reflected, in theory, the time and resources required to determine the relevant law and analyze the likely outcome.

Big data turns this approach on its head. Rather than assuming that rules dictate outcomes as the basis for making specific predictions, big data looks for patterns and correlations. For example, historical litigation data in the aggregate might reveal a judge's tendency to grant or deny certain types of pretrial motions, an opponent's historical avoidance of expert witnesses, or a party's typical timing for settlements. These details may be more relevant for a client or a lawyer than the published court opinions in prior cases that ran the full course of litigation.

Coasean bargaining between attorneys traditionally took place in the shadow of the trial, a result of the fact that litigants alter their pretrial behavior, including their willingness to negotiate a settlement, based on their perception of the likely outcome at trial and the anticipated cost to get there.[16] Traditionally, lawyers have formed their perception of the relative strength of their clients' cases based on case precedents, their prior experience handling similar cases, and interactions with the presiding judge and opposing counsel. Today, however, the availability of tools for leveraging legal data is increasingly moving bargaining into the shadow of the trends and patterns observable in aggregated litigation data.

What exactly does this mean for the legal profession? A traditional lawyer with many years of experience would acquire awareness of some relevant patterns, tendencies, and habits—the fact that certain judges were always severe; certain lawyers were more affable than others, even

---

16. *See generally* Robert H. Mnookin & Lewis Kornhauser, *Bargaining in the Shadow of the Law: The Case of Divorce*, 88 YALE L.J. 950, 950 (1979) (discussing "the impact of the legal system on negotiations and bargaining that occur *outside* the courtroom" in divorce cases and "the primary function of contemporary divorce law not as imposing order from above, but rather as providing a framework within which divorcing couples can themselves determine their post-dissolution rights and responsibilities").

as opposing counsel; certain parties were unusually litigious; certain courthouses were painfully slow, for example. Such experiential knowledge supplemented the lawyer's knowledge of the legal rules, and these personal factors could sometimes matter more than the legal rules themselves. Seasoned lawyers practiced not just in the shadow of the trial, but also in the shadow of the traits and tendencies of everyone involved.

Ironically, these intangible aspects of traditional lawyering are the most analogous to lawyering in the shadow of big data. The new technology, however, makes available within minutes insights that once might have taken thirty years of practice experience to acquire. The latest legal-data services available to practitioners yield accurate statistical information about the preferences, persistence, and pace of the judges, parties, and lawyers involved in a matter. Big (legal) data not only compresses the time component of such knowledge from decades to minutes, but it also dramatically expands the breadth of potential observations.[17] That is to say, a lawyer can now "profile" complete strangers from other jurisdictions for factors the lawyer has never before encountered in his limited circle of acquaintances.

At the outset, one should dispense with the cartoonish image of big data as a mechanical arbitrator or adjudicator that will simply tell lawyers which party will win or lose a case, eliminating the need to do any research or give the matter much thought. To the contrary, big data potentially makes the lawyer's job much more complicated and nuanced. Working in the shadow of data poses more ambiguity, both normative and positive, than working in the shadow of trials, which, in turn, occur in the shadow of rules. In fact, big data poses a new set of hazards or perils for its users. The line dividing successful from unsuccessful attorneys will shift from those skilled in "reading" the other parties to a lawsuit in the traditional sense of the word toward those familiar with the capabilities and limitations of lawyering with the aid of big data. Specifically, the lawyers of tomorrow will know when patterns in litigation data make predictions, when predictions run the risk of becoming self-negating or self-fulfilling, and when to reduce their exposure to the systemic risks that will inevitably result from more attorneys beginning to practice law overshadowed by data.

This Article also proposes to move beyond the tired clichés about newfangled technologies eliminating jobs. It is a truism that automation replaces the workers who once manually performed the newly automated

---

17. *See, e.g.*, Matthew Chin, *New Data Compression Method Reduces Big-Data Bottleneck; Outperforms, Enhances JPEG: New Discovery Is Rooted in Physics and the Arts*, UCLA NEWSROOM (Dec. 18, 2003), http://newsroom.ucla.edu/releases/ucla-research-team-invents-new-249693.

task.[18] These authors believe, like many economists, that automation eventually shifts the workforce elsewhere, to other nonautomated tasks, rather than demobilizing any section of the workforce permanently. Some legal tasks, like document review, are becoming increasingly automated, meaning associate attorneys are in less demand for such tasks;[19] but these authors believe that demand for associates who can handle data-driven lawyering will continue to increase as the legal profession catches up to the data-centric approach to information utilized in other industries.

The thrust of this Article is that lawyering in the shadow of data is qualitatively different than the technology-enhanced lawyering of previous decades. This Article attempts to take a more theoretical approach than previous literature in this area in order to paint a picture of what big data really means for lawyers and judges, and it aims to debunk some common misperceptions along the way. It includes a description of the newest legal-technology startups and some predictions about anticipated market entrants in the near future. In addition, this Article defines—for the first time—some of the hazards inherent in basing legal decisions on historical data that other commentators have largely overlooked. Some of these hazards are normative, but most of them are descriptive and theoretical. The intent of these authors is not to discourage the use of big data, but rather to help the legal profession understand the advantages and pitfalls of lawyering in the shadow of such data.

The ensuing discussion proceeds along a standard roadmap: Part I will familiarize readers with the very recent technological upheaval in the American legal profession and will discuss the emergence of the "law-as-data" movement that has resulted. This discussion includes a brief overview of PACER and the commercial alternatives. Part II then delves into the nature of legal "predictions" that big data will enable and attempts to define the value and limitations of such data-driven predictions. This Part contains the crux of this Article's argument: that

---

18. Some have argued that the legal profession historically has been hesitant to embrace any new technologies because reduced hours—through automation—would reduce billable revenue. *See, e.g.*, William G. Ross, *The Ethics of Hourly Billing by Attorneys*, 44 RUTGERS L. REV. 1, 30 (1991) ("A more troubling reason for the widespread failure of attorneys to use word processors may be that managers of law firms fear that the time saved by use of computers would result in a net reduction of billable hours.").

19. *See, e.g.*, Ed Poll, *Technology Investments and Two-Handed Decision Making: How to Determine ROI When Considering Hardware and Software Upgrades*, 30 NO. 4 LEGAL MGMT. 52, 54 (2011) ("E-discovery software can analyze documents required for litigation discovery in a fraction of the time for a fraction of the cost when compared to using lawyers for the task."); John Markoff, *Armies of Expensive Lawyers, Replaced by Cheaper Software*, N.Y. TIMES (Mar. 4, 2011), http://www.nytimes.com/2011/03/05/science/05legal.html ("Now, thanks to advances in artificial intelligence, 'e-discovery' software can analyze documents in a fraction of the time for a fraction of the cost.").

big data significantly enhances Coasean bargaining by lowering information costs, which traditionally formed the most easily identifiable transaction costs for purposes of Coasean analysis of legal transactions and disputes. Part III initiates a long-overdue discussion about the theoretical problems with predictions and forecasts, and about lingering uncertainty in the legal arena. An important Section in Part III addresses the special problem of "Black Swans"—unforeseeable systemic risks that creep in as information-based decision-making allows tighter integration and coordination of independent players in the system. This Article concludes with a concise recap of the main points and suggestions of areas for further research.

I. BACKGROUND

This Part discusses the very recent technological upheaval in the American legal profession, as well as the emergence of the "law-as-data" movement that has resulted. This discussion includes a brief overview of PACER as a means of accessing legal data and the commercial alternatives thereto.

A. *Intuition is Overrated*

Lawyers, like accountants, doctors, and any other professionals, are human. Human nature tends toward clouded judgment from the interference of cognitive biases and emotion in the decision-making process.[20] Lawyers are often unaware of the subtle saboteurs of rationality that have infiltrated their subconscious thought.[21] Their range of experiences and finite capacity to retain information also inherently limit practitioners' thinking.[22] The ability to make sound judgments takes years of practice to develop and a lifetime to master.

---

20. *See generally* HERBERT A. SIMON, ADMINISTRATIVE BEHAVIOR 81 (2d ed. 1957) (describing the limits of "objective rationality"); RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS 25–26 (2008) (explaining how cognitive biases and question framing skew decisions).

21. *See* JÜRGEN T. REHM & VOLKER GADENNE, INTUITIVE PREDICTIONS AND PROFESSIONAL FORECASTS: COGNITIVE PROCESSES AND SOCIAL CONSEQUENCES 73–74 (1990).

22. *See* Bibas, *supra* note 12, at 2519–20. The same is true, of course, for judges. *See* Richard A. Posner, *The Rise and Fall of Judicial Self-Restraint*, 100 CALIF. L. REV. 519, 553–54 (2012) ("Justices operate on limited information; because there are no sensible algorithmic methods of deciding difficult cases, most constitutional decisions have only weak claims to objective validity; the parts of the Constitution that generate litigation at the Supreme Court level are too old and general to be directive; the issues presented in constitutional cases tend to be both emotional and momentous and the decisions resolving them inescapably reflect the Justices' personal values, psychology, background, peer pressures, political anxieties, professional experiences, ideological inclinations, and other non-legalistic factors, often operating unconsciously . . . .").

Data, on the other hand, is indifferent. It is impartial. It can be comprehensive in scope where lawyers are limited in experience. Data does not care whether it is consistent with someone's preconceived notions or personal experiences, nor is it particularly concerned with intuition. Rather, data is simply a historical reflection of reality. It seems rather obvious, then, that aggregated data about relevant, past litigation should play a key role in informing the inherently subjective judgment of lawyers—a "reality check," so to speak.

Yet lawyers continue to rely mostly on their experience,[23] intuition,[24] and instinct.[25] This approach to legal decision-making is problematic for

---

23. *See* Brandon J. Harrison, *The Lawyer as a Professional Writer*, 62 ARK. L. REV. 725, 739–40 (2009) (encouraging lawyers to draw on their life experiences); Patrick E. Longan, *Bureaucratic Justice Meets ADR: The Emerging Role for Magistrates as Mediators*, 73 NEB. L. REV. 712, 726 (1994) (emphasizing the value of practitioner experience in mediation).

24. In 1993, the American Bar Association published the results of a study in which roughly 3000 lawyers were asked to take a Myers–Briggs Type Indicator test to measure the personality types prevalent in the profession. Larry Richard, *The Lawyer Types*, A.B.A. J., July 1993, at 74, 74. The study's results showed that although a mere thirty percent of the general U.S. population indicated that they relied primarily on their intuition rather than concrete facts, fifty-seven percent of lawyers were inclined to trust their intuition over concrete facts when making decisions. *Id.* at 76. This may be due in large part to the fact that lawyers often make judgment calls based on incomplete facts and perhaps are more accustomed to wrestling with ambiguity than the general population.

25. Edward S. Adams & Christian J. Lane, *Constructing a Jury That Is Both Impartial and Representative: Utilizing Cumulative Voting in Jury Selection*, 73 N.Y.U. L. REV. 703, 706 (1998) ("Attorneys rely on instinct and experience . . . ."); Sally E. Anderson, *Special Considerations for Sole and Small Firm Practitioners*, 32 WYO. LAW. 50, 56 (2009) ("Most importantly, a lawyer must trust his or her instincts."); Robert A. Caplen, *When* Batson *Met* Grutter*: Exploring the Ramifications of the Supreme Court's Diversity Pronouncements Within the Computerized Jury Selection Paradigm*, 10 U. PA. J. CONST. L. 65, 70 (2007) ("Lawyers often rely upon hunches, instinct, and educated guesswork, none of which are exact." (footnote omitted) (internal quotation marks omitted)); William R. Dailey, *Who is the Attorney General's Client?*, 87 NOTRE DAME L. REV. 1113, 1125 (2012) ("Presumably, a lawyer must still rely on her own instincts about how far to follow precedent, and so an account of precedent is bound to be part of any given lawyer's own 'best view' of the law, and the subjective element does not appear to be eliminated by requiring a 'due' respect for precedent."); Alan C. Eidsness & Lisa T. Spencer, *Confronting Ethical Issues in Practice: The Trial Lawyer's Dilemma*, 45 FAM. L.Q. 21, 27 (2011) ("[L]awyers should trust their instincts . . . ."); Sandra Anderson Garcia & Robert Batey, *The Roles of Counsel for the Parent in Child Dependency Proceeding*, 22 GA. L. REV. 1079, 1099 (1988) ("To be sure, much of case selection must rely on attorney instincts and experience . . . ."); Gerry Riskin, *Stealth Discrimination: A Model for Choosing and Managing Your Leaders*, 20 OF COUNS. 16, 18 (2001) ("My advice is to forget the labored refinements of lawyering. Trust your visceral instincts. Go with your hunches."); William H. Simon, *Role Differentiation and Lawyers' Ethics: A Critique of Some Academic Perspectives*, 23 GEO. J. LEGAL ETHICS 987, 1000 (2010) ("In locating the 'bounds of the law' for the purpose of counseling their clients, lawyers instinctively rely on principles, policies, and informal norms."); Dennis v. United States, 339 U.S. 162, 173 (1950) (Jackson, J., concurring) ("The common instinct of men for fair dealing and the experience of trial lawyers alike reject this holding.").

several reasons. Intuition takes years to develop.[26] To make matters worse, studies have shown that lawyers are generally overconfident when assessing their likelihood of success,[27] and that the accuracy with which lawyers can predict outcomes does not improve with experience.[28]

Considering what is at stake, it seems imprudent to rely on experience, intuition, or instinct alone in predicting the path of the law. As the next Section will explain, however, lawyers are not entirely to blame. Only recently have economic, technological, and political forces come together to form the perfect storm for introducing the legal profession to a new, data-driven approach to practicing law.

B. *Innovation in the Legal Industry*

This Article is set against the backdrop of the economic turmoil that has rocked the legal profession in recent years, the disruptive technology that is reshaping the decision-making processes in other industries, and the growing body of free and easily accessible legal data that now exists. For nearly a century, the practice of law remained partly sheltered from the winds of technological change that have swept through other industries in the recent past. That all changed in 2008 as some of the nation's largest white-shoe law firms collapsed under the crushing weight of overhead costs that clients could no longer afford to subsidize.[29] Many

---

26. *See generally* MARC J. DOLLINGER, ENTREPRENEURSHIP: STRATEGIES & RESOURCES 34 (4th ed. 2008); Elisha Babad, Michael Hills, & Michael O'Driscoll, *Factors Influencing Wishful Thinking and Predictions of Election Outcomes*, 13 BASIC & APPLIED SOC. PSYCHOL. 461, 474 (1992); Ziva Kunda, *The Case for Motivated Reasoning*, 108 PSYCHOL. BULL. 480, 487 (1990); Thomas A. Langens, *Regulatory Focus and Illusions of Control*, 33 PERSONALITY & SOC. PSYCHOL. BULL. 226, 226 (2007); Neil D. Weinstein, *Reducing Unrealistic Optimism About Illness Susceptibility*, 2 HEALTH PSYCHOL. 11, 12 (1983); Neil D. Weinstein, *Unrealistic Optimism About Susceptibility to Health Problems*, 5 J. BEHAV. MED. 441, 443 (1982); Neil D. Weinstein, *Unrealistic Optimism About Future Life Events*, 39 J. PERSONALITY & SOC. PSYCHOL. 806, 807 (1980).

27. Some have suggested that this is due in part to the fact that overconfidence is a necessary trait to attract and retain clients. Elizabeth F. Loftus & Willem A. Wagenaar, *Lawyers' Predictions of Success*, 28 JURIMETRICS J. 437, 450 (1988); *see also* William H. Simon, *Ethical Discretion in Lawyering*, 101 HARV. L. REV. 1083, 1139 (1988); Jack W. Brehm, *Postdecision Changes in the Desirability of Alternatives*, 52 J. ABNORMAL & SOC. PSYCHOL. 384, 389 (1956) (examining cognitive dissonance in the realm of alternative choices).

28. *See* Jane Goodman-Delahunty et al., *Insightful or Wishful: Lawyers' Ability to Predict Case Outcomes*, 16 PSYCHOL. PUB. POL'Y & L. 133, 133 (2010).

29. *See* Larry E. Ribstein, *The Death of Big Law*, 2010 WIS. L. REV. 749, 751–52; Steven Pearlstein, *Why Howrey Law Firm Could Not Hold It Together*, WASH. POST (Mar. 19, 2011), http://articles.washingtonpost.com/2011-03-19/business/35207602_1_law-firm-howrey-simon-robert-ruyak (observing that that rising overhead costs and fee cuts demanded by clients precipitated the 2011 collapse of the Howrey law firm, which "appeared to the outside world to be a thriving global enterprise with more than 700 lawyers" just two years earlier); James B. Stewart, *Dewey's Fall Underscores Law Firms' New Reality*, N.Y. TIMES (May 4, 2012),

more law firms have experienced dramatic downsizing in recent years.[30] With more empty chairs at the office and tighter lines of credit at the bank, law firms searched desperately for a cheaper alternative to rehiring costly junior associates when legal grunt work—document review— picked back up.[31]

Many law firms were quick to embrace e-discovery software, which automated the process of reviewing documents for relevancy with greater accuracy than human reviewers at a fraction of the cost.[32] As a result, the e-discovery-services industry sprang up virtually overnight, raking in revenue of approximately $2.8 billion in 2009 alone.[33] It was not long before the *New York Times* picked up on the significance of this shift toward automated legal services, heralding a day that many in the legal profession thought would never come, with a headline that read, "Armies of Expensive Lawyers, Replaced by Cheaper Software."[34]

---

http://www.nytimes.com/2012/05/05/business/deweys-collapse-underscores-a-new-reality-for-law-firms-common-sense.html.

30. *See* John B. Attanasio, *Facing Challenging Economic Times Head On*, 74 Tex. B.J. 138, 139 (2011); Richard W. Bourne, *The Coming Crash in Legal Education: How We Got Here, and Where We Go Now*, 45 Creighton L. Rev. 651, 658 (2012) ("From the beginning of January 2008 through November 2009, one website counted 14,940 layoffs from 'major law firms,' including 5,820 attorneys and 9,120 'staff' (comprising staff attorneys and contract attorneys as well as non-lawyer personnel), and the number of attorneys in the National Law Journal . . . 250 decreased 4.3% in 2009 and another 1.1% in 2010."); Ted Enarson, *Lateral Screening: Why Your State Should Not Adopt Amended Model Rule of Professional Conduct 1.10*, 37 J. Legal Prof. 1, 9 n.40 (2012); Luz E. Herrera, *Training Lawyer-Entrepreneurs*, 89 Denv. U. L. Rev. 887, 887 n.1 (2012); A. Benjamin Spencer, *The Law School Critique in Historical Perspective*, 69 Wash. & Lee L. Rev. 1949, 1953 (2012); Nathan Koppel, *Recession Batters Law Firms, Triggering Layoffs, Closings*, Wall St. J. (Jan. 26, 2009, 11:59 PM), http://www.wsj.com/articles/SB123292954232713979; Karen Sloan, *Another Week, Another Round of Firm Layoffs: In First Two Months of the Year, Legal Sector Loses 5,500 Jobs*, Nat'l L.J. (Mar. 16, 2009), http://www.nationallawjournal.com/id=1202429069191/Another-week-another-round-of-firm-layoffs (LEXIS subscription required) (describing staff and lawyer layoffs at several law firms).

31. *See generally* Manjoo, *supra* note 2 ("E-discovery software has been similarly revolutionary. These systems can mine huge volumes of material (like all the email correspondence in a civil suit) for damning evidence. The simplest software looks for specific keywords, but more sophisticated systems can detect patterns of behavior that might interest lawyers. This was the sort of work that once consumed the lives of first-year associates; now computers do it faster, at lower cost, and with about as much success as humans.").

32. *See supra* note 19 and accompanying text.

33. Arin Greenwood, *A New View, Part 2: E-Discovery Changes Have Some Seeing a Career in Document Review*, A.B.A. J., Oct. 2011, at 27, 27–28.

34. Markoff, *supra* note 19.

C. *Law as Data*

This Section discusses the emergence of the "law-as-data" movement. It starts with a look at the rise of data in other sectors of industry, and then discusses the way the legal sector interacts with this rising tide of information. Finally, it examines reasons and means for the legal profession's close relationship with this data in litigation.

1. Big Data in Disparate Domains

While technological innovation has crept into certain discrete facets of the legal profession, it has dramatically improved other sectors of the economy. Technology giants like Google, Amazon, and Facebook have demonstrated the great value of making the world's information more useful and accessible. Even traditional brick-and-mortar industry leaders like Wal-Mart and General Electric now heavily rely on sophisticated computer networks and software to collect, store, distill, and act on the deluge of data that pours into their servers daily.[35] The ability to crunch big data in this manner largely stems from two recent developments: the abundance of data that now exists and the emergence of software that makes it more accessible and useful.[36]

So what exactly is big data? For some, it is a tired cliché, nothing more than "[a] meme and a marketing term."[37] For others, it is like gold[38] or the new oil—"valuable, but if unrefined it cannot really be used."[39] National Public Radio likened big data to dust: using a cell phone, browsing the web, buying groceries, connecting with friends on Facebook, driving with a toll tag—each day "[w]e kick up clouds of it wherever we go."[40] The McKinsey Global Institute defined the term more formally as "large pools of data that can now be captured, communicated, aggregated, stored, and analyzed."[41] However one describes or defines big data, there is no denying that its time has come.[42] "[T]he world

---

35. *See The Data Deluge*, ECONOMIST (Feb. 25, 2010), http://www.economist.com/node/15579717.

36. AYRES, *supra* note 5, at 138; Katz, *supra* note 2, at 913–14.

37. Steve Lohr, *The Age of Big Data*, N.Y. TIMES (Feb. 11, 2012), http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html.

38. *Id*.

39. Michael Palmer, *Data is the New Oil*, ANA MARKETING MAESTROS (Nov. 3, 2006, 5:43 AM), http://ana.blogs.com/maestros/2006/11/data_is_the_new.html.

40. Geoff Nunberg, *Forget YOLO: Why 'Big Data' Should Be the Word of the Year*, NPR (Dec. 20, 2012, 10:58 AM), http://www.npr.org/2012/12/20/167702665/geoff-nunbergs-word-of-the-year-big-data.

41. JAMES MANYIKA ET AL., MCKINSEY GLOBAL INST., BIG DATA: THE NEXT FRONTIER FOR INNOVATION, COMPETITION, AND PRODUCTIVITY, at iv (2011), *available at* http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx.

42. *See* Lohr, *supra* note 37.

contains an unimaginably vast amount of digital information which is getting ever vaster ever more rapidly."[43] Thanks to a phenomenon known as Kryder's Law, storage capacity has dramatically increased in step with the explosion of data, while its cost continues to fall.[44]

The abundance of data and the increasing capacity of computers to store it are only half of the equation. The other half consists of computers that are now capable of scouring big data for patterns that might help human analysts solve problems and answer questions. Baseball teams crunch obscure statistics to spot undervalued players.[45] President Barack Obama's 2012 reelection campaign identified likely donors and undecided voters by combing through mountains of polling data.[46] Commercial industries as diverse as healthcare,[47] national defense,[48]

---

43. *Data, Data Everywhere*, ECONOMIST (Feb. 25, 2010), http://www.economist.com/node/15557443.

44. Katz, *supra* note 2, at 916.

45. LEWIS, *supra* note 7, at xiv.

46. Michael Scherer, *Inside the Secret World of the Data Crunchers Who Helped Obama Win*, TIME (Nov. 7, 2012), http://swampland.time.com/2012/11/07/inside-the-secret-world-of-quants-and-data-crunchers-who-helped-obama-win/2/. The article concluded:

> That data-driven decisionmaking played a huge role in creating a second term for the 44th President and will be one of the more closely studied elements of the 2012 cycle. It's another sign that the role of the campaign pros in Washington who make decisions on hunches and experience is rapidly dwindling, being replaced by the work of quants and computer coders who can crack massive data sets for insight. As one official put it, the time of "guys sitting in a back room smoking cigars, saying 'We always buy *60 Minutes*'" is over. In politics, the era of big data has arrived.

Id.

47. *See generally* PETER GROVES ET AL., CTR. FOR U.S. HEALTH REFORM BUS. TECH. OFFICE, THE "BIG DATA" REVOLUTION IN HEALTHCARE: ACCELERATING VALUE AND INNOVATION 1 (Jan. 2013), *available at* http://www.mckinsey.com/~/media/mckinsey/dotcom/client_service/healthcare%20systems%20and%20services/pdfs/the_big_data_revolution_in_healthcare.ashx.

48. Chris Young, *Military Intelligence Redefined: Big Data in the Battlefield*, FORBES (Mar. 12, 2012, 12:53 PM), http://www.forbes.com/sites/techonomy/2012/03/12/military-intelligence-redefined-big-data-in-the-battlefield/.

fraud detection,[49] travel,[50] online dating,[51] telecommunications,[52] oil and gas,[53] and banking[54] have enjoyed similar benefits by figuring out how to capture and glean insight from the deluge of domain-specific data that now exists.

The legal profession has been noticeably absent from the parade of progress that has led to valuable breakthroughs in other industries. It seems anomalous that "a singularly information-dependent profession" like the legal profession would arrive late to the age of big data.[55] Some commentators have suggested that lawyers have resisted technological progress for fear that efficiency gains might decrease the number of billable hours for which they may charge their clients.[56] While some lawyers may indeed harbor such unsavory motives for resisting innovation, this Article suggests that the best attorneys welcome new solutions that let them perform at an even higher level for a greater number of clients. The more plausible explanation for the lack of big-data solutions in the legal-research space is the historical inaccessibility of legal data, as discussed below.

---

49. *See The Data Deluge*, *supra* note 35 ("Credit-card companies monitor every purchase and can identify fraudulent ones with a high degree of accuracy, using rules derived by crunching through billions of transactions. Stolen credit cards are more likely to be used to buy hard liquor than wine, for example, because it is easier to fence. Insurance firms are also good at combining clues to spot suspicious claims: fraudulent claims are more likely to be made on a Monday than a Tuesday, since policyholders who stage accidents tend to assemble friends as false witnesses over the weekend. By combining many such rules, it is possible to work out which cards are likeliest to have been stolen, and which claims are dodgy.").

50. *Data, Data Everywhere*, *supra* note 43 ("Microsoft's search engine Bing, can advise customers whether to buy an airline ticket now or wait for the price to come down by examining 225 billion flight and price records.").

51. Lohr, *supra* note 37 ("Online dating services, like Match.com, constantly sift through their Web listings of personal characteristics, reactions and communications to improve the algorithms for matching men and women on dates."); John Tierney, *Hitting It Off, Thanks to Algorithms of Love*, N.Y. TIMES (Jan. 29, 2008), http://www.nytimes.com/2008/01/29/science/29tier.html.

52. *The Data Deluge*, *supra* note 35 ("Mobile-phone operators, meanwhile, analyse subscribers' calling patterns to determine, for example, whether most of their frequent contacts are on a rival network. If that rival network is offering an attractive promotion that might cause the subscriber to defect, he or she can then be offered an incentive to stay.").

53. *See id.* ("The oil industry uses supercomputers to trawl seismic data before drilling wells.").

54. *Data, Data Everywhere*, *supra* note 43 ("Personal-finance websites and banks are aggregating their customer data to show up macroeconomic trends, which may develop into ancillary businesses in their own right.").

55. *See* Richard L. Marcus, *The Impact of Computers on the Legal Profession: Evolution or Revolution?*, 102 NW. U. L. REV. 1827, 1829 (2008).

56. *See, e.g.*, Ross, *supra* note 18, at 30.

2. Viewing the Law from a New Perspective

There are many lenses through which to view the law. Justice Holmes viewed the law through the eyes of a "bad man."[57] "For many observers of the American legal system, law is what judges write in appellate opinions."[58] Having entered the age of big data, a growing number of scholars are now urging academics and practitioners alike to view the law as data. For instance, Professors Daniel Katz[59] and Josh Blackman[60] routinely write and speak about the value of aggregating bulk legal data and then analyzing it using modern computing. This "law-as-data" movement stems from the legal academy's growing interest in legal empiricism, which, until recently, remained largely relegated to the realm of political science.[61]

To date, the vast majority of empirical studies of the law—or more precisely, litigation—have focused on the U.S. Supreme Court.[62] This is understandable given the High Court's position at the top of the Judicial Branch, its greater visibility compared to lower courts, and its smaller, more manageable pool of cases. A handful of scholars have, however, stressed the importance of empirically analyzing litigation data generated at the district-court level.[63]

---

57. Holmes, *supra* note 1, at 459.

58. David A. Hoffman et al., *Docketology, District Courts, and Doctrine*, 85 WASH. U. L. REV. 681, 683 (2007).

59. *See generally* Katz, *supra* note 2, at 910–11 (discussing the significant impact innovations like e-discovery will have on the legal market).

60. Josh Blackman et al., *Fantasy SCOTUS: Crowdsourcing a Prediction Market for the Supreme Court*, 10 NW. J. TECH. & INTELL. PROP. 125, 165 (2012) (proposing to use "a bot to crawl through all of the filings in PACER . . . and develop a comprehensive database of all aspects of how each court works").

61. *See generally* Gerald N. Rosenberg, *Across the Great Divide (Between Law and Political Science)*, 3 GREEN BAG 2d 267, 268 (2000) (noting law professors' inabilities to contribute to empirical political science scholarship due to lack of training).

62. *See, e.g.*, Lee Petherbridge & David L. Schwartz, *An Empirical Assessment of the Supreme Court's Use of Legal Scholarship*, 106 NW. U. L. REV. 995, 998 (2012); Theodore W. Ruger et al., *The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking*, 104 COLUM. L. REV. 1150, 1151 (2004); Louis J. Sirico Jr. & Jeffrey B. Margulies, *The Citing of Law Reviews by the Supreme Court: An Empirical Study*, 34 UCLA L. REV. 131, 135 (1986); David C. Thompson & Melanie F. Wachtell, *An Empirical Analysis of Supreme Court Certiorari Petition Procedures: The Call for Response and the Call for the Views of the Solicitor General*, 16 GEO. MASON L. REV. 237, 240 (2009); Nicholas S. Zeppos, *The Use of Authority in Statutory Interpretation: An Empirical Analysis*, 70 TEX. L. REV. 1073, 1088 (1992).

63. *See, e.g.*, Hoffman et al., *supra* note 58, at 684 (describing the benefits of conducting empirical research of district courts using the full case records); Pauline T. Kim et al., *How Should We Study District Judge Decision-Making?*, 29 WASH. U. J.L. & POL'Y 83, 103 (2009) ("PACER offers a significant data source for more accurately capturing and understanding the activity of the district courts[than do published opinions]."); Margo Schlanger & Denise Lieberman, *Using Court Records for Research, Teaching, and Policymaking: The Civil Rights Litigation*

For the purposes of this Article, "litigation data" means: (1) the basic information about a given case, such as the names of the parties, attorneys, nature of the suit, and other characteristics about the case (otherwise referred to as case metadata); (2) information reflected on the electronic docket sheet for a given case; and (3) the electronic court records associated with a given case.[64] When combined, these sources paint a detailed portrait of litigation on the front lines of the federal Judiciary. When aggregated on a large scale and then filtered using software, patterns and trends begin to surface that shed light on how parties, attorneys, and judges typically behave during the litigation process.[65]

Case metadata becomes even more valuable when combined with information scraped from case docket sheets and then classified using a machine-learning algorithm.[66] Docket sheets are ideally suited for studying the ebb and flow of litigation because they consist of a sequentially ordered timeline of every event that occurs over the life of a lawsuit.[67] When a lawyer electronically files something or a judge communicates in some fashion with the parties, the court clerk will create a new docket entry that consists of a brief description of the event, the date of that event, and a hyperlink to any documents associated with that event.[68]

The text passages in docket entries lend themselves to large-scale analysis because they are far easier to automatically parse and classify with machine-learning algorithms than court records (a key component of software-enabled, trend-based lawyering), as the text is typically short and formulaic. Docket sheets are also an important source of information about district courts because the vast majority of litigation events do not

---

Clearinghouse, 75 UMKC L. REV. 155, 163, 168 (2006) ("In short, for anyone who hopes to understand litigation—one specific litigation or an entire field of litigation—there is no substitute for court records.").

64. This Article limits its analysis to litigation data collected from district courts at the federal level, although many of the principles discussed herein apply equally to litigation data at the state level.

65. *See generally* Hoffman et al., *supra* note 58, at 683–85 (describing the benefits of conducting empirical research about district courts using the full case records); Schlanger & Lieberman, *supra* note 63, at 168 (underlining the importance of court records to understanding litigation).

66. For the purposes of this Article, the authors collaborated with an expert in scraping and machine learning to develop a proprietary algorithm capable of automatically reading case docket sheets, identifying summary judgment events, and classifying the outcome, movants, nonmovants, and other metadata that the authors may discuss at greater length in a subsequent article.

67. *See* Kim et al., *supra* note 63, at 107–08.

68. Brian N. Lizotte, *Publish or Perish: The Electronic Availability of Summary Judgments by Eight District Courts*, 2007 WIS. L. REV. 107, 126.

produce written opinions.[69] As one commentator has pointed out, "a district judge may make many decisions of varying types at different points in time in a single case," as most decisions "are not accompanied by written reasons and therefore are recorded on the docket sheet or in a brief order as simply a decision to 'grant' or 'deny' a particular motion."[70] Decisions that do not produce written opinions, however, can sometimes dramatically alter the cost and ultimate outcome of litigation.[71]

Readers are probably most familiar with the third and final category of litigation data discussed in this Article—court records. Court records tell the story of a given case. To initiate a lawsuit in district court, a plaintiff must file a complaint[72] that sets forth a valid legal basis for relief and factual allegations that would, if true, justify such relief.[73] A defendant will then respond to the complaint by filing an answer,[74] motion to dismiss,[75] or any number of other legal instruments.[76] After reaching a final decision in a given case, a judge may issue an order, ruling, or judgment accompanied by a written opinion in which the judge will articulate her findings of fact and conclusions of law in support of her decision.

3. Public Access to Legal Data

The previous Section answered the "What" question: that is, "What type of data may lead to a better understanding of litigation?" Before moving on to the "Where" question (as in, "Where can this data be found?"), it is important to first take a step back and ask, "Why should this data be accessible for empirical analysis?" The answer to this question requires a brief journey back in time.

Since the age of the Pharaohs, men and women have conditioned their willingness to submit to the judgment of another on the guarantee that such decisions would be carried out in public. Inscribed inside the tomb of an Egyptian official who reported to King Tuthmosis III were these instructions:

> Lo, whenever an administrator bears cases let there be publicity . . . and . . . let water and air report all that he may do. Lo, then his conduct is by no means unperceived. If he

---

69. *See* Kim et al., *supra* note 63, at 98.

70. *See id.* (discussing opinion bias in empirical studies of district courts).

71. *Id.* at 94, 106–08 (describing the ways in which discovery decisions and nondispositive rulings can influence case settlements).

72. FED. R. CIV. P. 3.

73. FED. R. CIV. P. 8(a).

74. FED. R. CIV. P. 12(a).

75. FED. R. CIV. P. 12(b).

76. *See, e.g.*, FED. R. CIV. P. 12(c), (e).

> does anything [unseemly] and he is to blame . . . he is not to
> be reinstalled . . . on the authority of an acting official but
> men shall learn of it . . . on the authority of his (proper)
> judge . . . .[77]

The strong preference for transparency in U.S. courts traces its roots back to Medieval Europe. In the mid-thirteenth century, lawyers began keeping records of their appearances before the English court of common pleas.[78] At the end of their career, lawyers would hand their records down to their apprentice, who would then carry on the tradition.[79] Lawyers quickly realized the value of keeping detailed records of proceedings, often citing them to show that their proffered position was consistent with the judge's previous rulings or that the position advanced by their adversary was inconsistent with previous rulings.[80] This practice developed into "the concept of precedent as a means towards achieving uniformity and consistency in pleading and in judicial decision-making."[81]

Lawyers and judges were not the only ones to benefit from the accessibility of public records. English peasants used court records to prove their entitlement to develop land under the feudal system.[82] Before long, court records were considered to be a "public resource."[83] The importance of transparency was also not lost on the colonists who established the first courts in the United States. For example, the 1677 Concessions and Agreements of West New Jersey formally recognized that "inhabitants of the said Province may freely come into, and attend the said courts, and hear and be present, at all or any such tryals [sic] as shall be there had or passed, that justice may not be done in a corner nor in any covert manner."[84] The colonial government of Pennsylvania similarly declared, in 1682, that "all courts shall be open."[85] Early settlers

---

77. NORMAN DE GARIS DAVIES, THE METROPOLITAN MUSEUM OF ART EGYPTIAN EXPEDITION, THE TOMB OF REKH-MI-RE 86–87 (Ludlow Bull ed., 1944) (footnote omitted) (offering possible translations of the hieroglyphics from the tomb of Rekh-mi-Re, Vizier to Tuthmosis III, King of Egypt, XVIII Dynasty (1580–1321 B.C.)).

78. Anthony Musson, *Law and Text: Legal Authority and Judicial Accessibility in the Late Middle Ages*, *in* THE USES OF SCRIPT AND PRINT, 1300–1700, at 103 (Julia Crick & Alexandra Walsham eds., 2004).

79. *See, e.g.*, *id.* at 104.

80. *Id.* at 103.

81. *Id.*

82. *See id.* at 104.

83. *Id.*

84. CONCESSIONS AND AGREEMENTS OF WEST NEW JERSEY, 1677, *reprinted in* SOURCES OF OUR LIBERTIES 180, 188 (Richard L. Perry ed., 1959).

85. FRAME OF GOVERNMENT OF PENNSYLVANIA, 1682, *reprinted in* SOURCES OF OUR LIBERTIES 209, 217 (Richard L. Perry ed., 1959).

"sensed from experience and observation that . . . the means used to achieve justice must have the support derived from public acceptance of both the process and its results."[86]

Today, federal courts carry on the nation's tradition of transparency by preserving the public's broad, common-law "right to inspect and copy public records and documents, including judicial records and documents."[87] This right of public access to court records applies to both civil and criminal actions, and includes both pretrial and trial proceedings.[88] Some courts have gone a step further by recognizing the right to access public records in certain cases as one that is protected by the First Amendment.[89]

---

86. Richmond Newspapers, Inc. v. Virginia, 448 U.S. 555, 571 (1980).

87. *See, e.g.*, Nixon v. Warner Commc'ns, Inc., 435 U.S. 589, 597 (1978); *Ex parte Drawbaugh*, 2 App. D.C. 404, 407–08 (1894) ("Such claims of right, and contests over them, are not the ordinary incidents of judicial proceeding; and any attempt to maintain secrecy, as to the records of the court, would seem to be inconsistent with the common understanding of what belongs to a public court of record, to which all persons have the right of access, and to its records, according to long established usage and practice.").

88. *See, e.g.*, Hartford Courant Co. v. Pellegrino, 380 F.3d 83, 93 (2d Cir. 2004) (protecting the public's right to inspect and copy docket sheets); Foltz v. State Farm Mut. Auto. Ins. Co., 331 F.3d 1122, 1134 (9th Cir. 2003) (noting that the common law right to public access extends "to pretrial documents filed in civil cases, including materials submitted in connection with motions for summary judgment"); Smith v. U.S. Dist. Ct. for So. Dist. of Ill., 956 F.2d 647, 649 (7th Cir. 1992) ("The federal common law right of access to judicial records and documents is well recognized."); FTC v. Standard Fin. Mgmt. Corp., 830 F.2d 404, 409 (1st Cir. 1987) ("[W]e rule that relevant documents which are submitted to, and accepted by, a court of competent jurisdiction in the course of adjudicatory proceedings, become documents to which the presumption of public access applies."); *In re* Cont'l Ill. Sec. Litig., 732 F.2d 1302, 1308–09 (7th Cir. 1984); United States v. Edwards, 672 F.2d 1289, 1294 (7th Cir. 1982).

89. *See* United States v. Ochoa-Vasquez, 428 F.3d 1015, 1029 (11th Cir. 2005); *Hartford Courant Co.*, 380 F.3d at 93; *In re* Providence Journal Co., 293 F.3d 1, 10–13 (1st Cir. 2002); Phoenix Newspapers, Inc. v. U.S. Dist. Court for Dist. of Ariz., 156 F.3d 940, 945, 948 (9th Cir. 1998) (concluding that the First Amendment "requires release of transcripts [of closed criminal proceedings] when the competing interests precipitating hearing closure are no longer viable"); United States v. Antar, 38 F.3d 1348, 1359–60 (3d Cir. 1994) ("[T]he right of access to voir dire examinations encompasses equally the live proceedings and the transcripts which document those proceedings. . . . It is access to the content of the *proceeding*—whether in person, or via some form of documentation—that matters."); Grove Fresh Distrib., Inc. v. Everfresh Juice Co., 24 F.3d 893, 897 (7th Cir. 1994) ("The First Amendment presumes that there is a right of access to proceedings and documents which have historically been open to the public and where the disclosure of which would serve a significant role in the functioning of the process in question." (internal quotation marks omitted)); Globe Newspaper Co. v. Pokaski, 868 F.2d 497, 505 (1st Cir. 1989) (finding that "after *Richmond Newspapers*, a blanket prohibition on the disclosure of records of closed criminal cases of the types at issue here implicates the First Amendment"); *In re* Search Warrant for Secretarial Area-Gunn, 855 F.2d 569, 573 (8th Cir. 1988) (stating that although "[t]he Supreme Court has not addressed the question whether the [F]irst [A]mendment right of public access extends to *documents*," it was "persuaded that the [F]irst [A]mendment right of public access does extend to the documents filed in support of search warrant applications");

Courts and scholars alike have identified a variety of benefits that flow from public access to court records, including: (1) courts that are more accountable to the people;[90] (2) courts that are more accountable to policymakers;[91] (3) increased public respect for and confidence in the court system;[92] (4) greater consistency in the law's application;[93] (5) increased public understanding of how laws are made and applied;[94] and (6) the "community therapeutic value" of openness that squelches the public's urge to take justice into its own hands.[95]

4. Sources of Litigation Data

This Subsection discusses various sources of litigation data. It starts with an overview of the system utilized to access federal court litigation data, and then proceeds to look at alternatives to that system.

a. PACER

Federal litigation data is available electronically to the public through Public Access to Court Electronic Records (PACER), an online repository for court records, docket sheets, and case metadata maintained

---

Anderson v. Cryovac, Inc., 805 F.2d 1, 5–7 (1st Cir. 1986) (discussing First Amendment cases in other circuits). *But see* N. Jersey Media Grp., Inc. v. Ashcroft, 308 F.3d 198, 200–02 (3d Cir. 2002) (ruling newspapers did not have right of access to civil immigration proceedings); Wilson v. Am. Motors Corp., 759 F.2d 1568, 1569–70 (11th Cir. 1985) (explaining that the Eleventh Circuit had not explicitly joined these constitutional holdings but had recognized a right of access to certain fundamental aspects of civil proceedings).

    90. *See Smith*, 956 F.2d at 650; *Nixon*, 435 U.S. at 598 (noting that public's right to access court records arises from "the citizen's desire to keep a watchful eye on the workings of public agencies"); *see also* CESARE BECCARIA, ON CRIMES AND PUNISHMENTS 22 (Henry Paolucci trans., Bobbs-Merrill Co. 1963) (1764) ("Let the verdicts and proofs of guilt be made public, so that opinion, which is, perhaps, the sole cement of society, may serve to restrain power and passions; so that the people may say, we are not slaves, and we are protected—a sentiment which inspires courage and which is the equivalent of a tribute to a sovereign who knows his own true interests."); Peter A. Winn, *Online Court Records: Balancing Judicial Accountability and Privacy in an Age of Electronic Information*, 79 WASH. L. REV. 307, 307 & n.3 (2004) [hereinafter Winn, *Online Court Records*].

    91. Lynn M. LoPucki, *Court-System Transparency*, 94 IOWA L. REV. 481, 495–97 (2009).

    92. Press-Enterprise Co. v. Super. Ct. of Cal., 464 U.S. 501, 508 (1984) ("Openness thus enhances both the basic fairness of the criminal trial and the appearance of fairness so essential to public confidence in the system.").

    93. LoPucki, *supra* note 91, at 507–08.

    94. Richmond Newspapers, Inc. v. Virginia, 448 U.S. 555, 572–73 (1980).

    95. *Id*. at 570–71; *see also* Brown & Williamson Tobacco Corp. v. F.T.C., 710 F.2d 1165, 1178 (6th Cir. 1983) ("The crucial prophylactic aspects of the administration of justice cannot function in the dark; no community catharsis can occur if justice is 'done in a corner [or] in any covert manner.'" (quoting *Richmond Newspapers*, 448 U.S. at 571)); Peter A. Winn, *Judicial Information Management in an Electronic Age: Old Standards, New Challenges*, 3 FED. CTS. L. REV. 135, 137 (2009) [hereinafter Winn, *Judicial Information Management*].

by the Administrative Office of the U.S. Courts.[96] Before there was PACER, there was paper. At that time, litigation data was practically obscure in that it took far more effort and expense to locate, copy, and aggregate case files when they were stashed away in filing cabinets buried in dimly lit basements of federal courthouses across the country.[97] Not surprisingly, empirical studies of district courts were relatively modest in size and few and far between during the infancy of the Internet.[98] Litigants had little fear that their personal affairs documented in court records would leak into the public domain, notwithstanding the fact that these documents were technically part of the public record.[99] District courts were also free to manage their dockets with minimal public scrutiny.[100]

That all changed with the creation of PACER. The Judiciary's switch from paper to electronic case filings made litigation data available to anyone with a computer, an internet connection, and a credit card.[101] Practical obscurity suddenly became a relic of the past. As discussed below, however, the Judiciary has recreated practical obscurity to a degree by erecting PACER's pay wall, which makes practice-focused empirical analysis of bulk litigation data prohibitively costly.

The federal Judiciary created PACER in 1990, "when an appropriations act authorized the federal judiciary to build a system furnishing remote access to court records, to be supported by funds generated by access fees."[102] The system's shift from keeping paper

---

96. Michel Ishakian & Ted Willmann, *About Us*, PACER, http://www.pacer.gov/about.html (last visited July 1, 2015).

97. Justice John Paul Stevens appears to have coined the U.S. Government's term "practical obscurity," which he used to refer to the inaccessibility of paper documents due to the inconvenience associated with acquiring access to them. U.S. Dep't of Justice v. Reporters Comm. for Freedom of the Press, 489 U.S. 749, 762–64 (1989). *See generally* Winn, *Judicial Information Management*, *supra* note 95, at 152 (describing the practical barriers to access associated with paper-based "practical obscurity").

98. *See, e.g.*, Bryant G. Garth, *Observations on an Uncomfortable Relationship: Civil Procedure and Empirical Research*, 49 ALA. L. REV. 103, 116 (1997) ("[W]e have relatively little reliable data over time on the civil justice system . . . ."); Marc Galanter et al., *How to Improve Civil Justice Policy*, 77 JUDICATURE 185 (1994) (calling for more empirical studies of the civil justice system); Susan W. Johnson & Donald R. Songer, *The Influence of Presidential Versus Home State Senatorial Preferences on the Policy Output of Judges on the United States District Courts*, 36 L. & SOC'Y REV. 657, 661 (2002) (noting the dearth of empirical studies focusing on the decision-making process of federal district courts).

99. *See* Winn, *Judicial Information Management*, *supra* note 95, at 138; Winn, *Online Court Records*, *supra* note 90, at 316–17.

100. Winn, *Judicial Information Management*, *supra* note 95, 168–70.

101. Peter W. Martin, *Online Access to Court Records—From Documents to Data, Particulars to Patterns*, 53 VILL. L. REV. 855, 860–61 (2008).

102. *Id.* at 860.

records to electronic records resulted in "an explosion of [the] use" of PACER as 20,028 user accounts in 1995 and 39,408 user accounts in 1999 dramatically increased to 270,000 user accounts by 2003.[103]

Recent observers heralded PACER's creation as "a huge improvement over the existing system of paper records,"[104] "tremendously ahead of its time,"[105] and "one of the great success stories of the federal Judiciary."[106] The enthusiasm with which users initially greeted PACER is not surprising considering the old method of inspecting court records, which entailed traveling to the courthouse where the desired documents were located, thumbing through boxes of dusty files, and photocopying them for a fee of fifty cents per page.[107] Despite the Judiciary's quantum leap from paper to pixels over two decades ago, PACER is beginning to show signs of aging. Critics now complain that the service imposes arbitrarily high access fees;[108] provides restrictive, inadequate search capabilities;[109] and has an outdated user interface that renders the data largely inaccessible.[110]

The most common complaint about PACER appears to be directed toward the service's access fees. Rather than subsidizing PACER's creation and maintenance with taxpayer dollars, Congress authorized the

---

103. *Id.* at 861.

104. Timothy B. Lee, *The Case Against PACER: Tearing Down the Courts' Paywall*, ARSTECHNICA (Apr. 9, 2009, 12:30 AM), http://arstechnica.com/tech-policy/2009/04/case-against-pacer/.

105. Steve Schultze, *PACER, RECAP, and the Movement to Free American Case Law*, VOXPOPULII (Feb. 3, 2011), http://blog.law.cornell.edu/voxpop/2011/02/03/pacer-recap-and-the-movement-to-free-american-case-law/.

106. *A Look at Electronic Public Access in the Federal Courts*, THIRD BRANCH, Aug. 2010, at 6, 6 (interviewing Judge J. Richard Leonard).

107. 28 U.S.C. § 1913(4) (Supp. V 1981).

108. *See, e.g.*, Greg Beato, *Tear Down This Paywall*, REASON (May 30, 2012), http://reason.com/archives/2012/05/30/tear-down-this-paywall; Timothy B. Lee, *Federal Courts Jack Up Fees for Online Access by 25 Percent*, ARSTECHNICA (Sept. 14, 2011, 4:05 PM), http://arstechnica.com/tech-policy/2011/09/federal-courts-jack-up-fees-for-online-access-by-25-percent/.

109. *See, e.g.*, Hillel Y. Levin, *Making the Law: Unpublication in the District Courts*, 53 VILL. L. REV. 973, 987 (2008) (describing how PACER's search engine restricts users' ability to search for unpublished legal opinions); Martin, *supra* note 101, at 869 (noting that PACER's search tools and interface "are sufficiently limited that they provide a market for commercial database offerings that consist of PACER data and a more sophisticated search engine"); Elizabeth Y. McCuskey, *Clarity and Clarification:* Grable *Federal Questions in the Eyes of Their Beholders*, 91 NEB. L. REV. 387, 443 (2012) (describing PACER's search capabilities as "lugubrious"); Lee F. Peoples, *The Citation of Wikipedia in Judicial Opinions*, 12 YALE J.L. & TECH. 1, 44 (2009) ("The PACER system has come under criticism recently for being difficult to use and expensive.").

110. *See* John Schwartz, *An Effort to Upgrade a Court Archive System to Free and Easy*, N.Y. TIMES (Feb. 12, 2009), http://www.nytimes.com/2009/02/13/us/13records.html.

Administrative Office of the U.S. Courts to charge reasonable user fees for access to electronic litigation data.[111] Therefore, to access court records through PACER, members of the public must register an account with the site and provide credit-card information.[112]

The service currently charges users ten cents per page to view and download PDFs of public court records.[113] The service also charges ten cents per search, regardless of whether the search results yield relevant cases.[114] While ten cents per page may seem like a trivial fee to most users, it makes compiling a sizeable dataset for empirical analysis prohibitively expensive.[115] Ten cents per page also seems far less reasonable when compared with the cost per printed page. One article pointed out that, under PACER's pricing scheme, "a weekday copy of the *New York Times* would go for $7 or $8, and Walter Isaacson's biography of Steve Jobs would cost $65.60."[116]

One of the most vocal critics of PACER's pay wall has been Carl Malamud, who has earned the reputation as being an "open government gadfly" thanks to his aggressive efforts to make digital copies of government records available to the public for free.[117] According to Malamud, "putting the nation's legal system behind a wall of cash and kludge separates the people from . . . the 'operating system for democracy.'"[118] So, when the Administrative Office of the U.S. Courts and the Government Printing Office announced that they would be providing users with free access to PACER records at seventeen libraries across the country as part of a pilot program,[119] Malamud saw the

---

111. *See* E-Government Act of 2002, Pub. L. No. 107-347, 116 Stat. 2899 (2002) (codified in scattered sections of 44 U.S.C.).

112. *PACER Registration*, PACER, http://www.pacer.gov/reg_pacer.html (last visited July 1, 2015).

113. PACER, ELECTRONIC PUBLIC ACCESS FEE SCHEDULE (Dec. 1, 2013), *available at* http://www.pacer.gov/documents/epa_feesched.pdf.

114. *Id.*

115. To PACER's credit, users get free access to PACER data as long as their accounts do not accrue "charges of more than $15.00 in a quarterly billing cycle." *Id.* The service also provides free access to judicial opinions, and courts may waive PACER fees for, among other user categories, individual researchers affiliated with a particular education institution. *Id.* That said, the steep PACER fees remain prohibitively costly to commercial ventures that might otherwise develop new tools around the data to improve the public's access to litigation information and understanding of the courts.

116. Beato, *supra* note 108.

117. *Id.*

118. Schwartz, *supra* note 110.

119. Press Release, Federal Depository Library Program, PACER Access (Sept. 26, 2008), http://www.fdlp.gov/index2.php?option=com_content&task=emailform&id=229&itemid=22 9 (suspending the pilot program to provide free access to PACER documents at seventeen libraries across the United States, "pending an evaluation").

opportunity that he had been waiting for and urged his supporters to download as many free court records through the PACER pilot program as possible.[120] He then asked his supporters to send the records to him so that he could make them available to the public for free on his website.[121]

Steve Schultze, Princeton University's Associate Director of Information Technology Policy at that time, felt compelled to join Malamud's "PACER liberation front," as he would later call it, after becoming frustrated with the system's inaccessibility.[122] Schultze realized, however, that PACER houses hundreds of millions of court records and that the number of records that he could manually download would amount to nothing more than a drop in the ocean.[123] So, instead, Schultze created a computer program that could download bulk records from PACER and automatically save them to a thumb drive.[124] Schultze gave the program to a fellow hacker named Aaron Swartz, who then proceeded to download around 2.7 million public court records from PACER for free from an account created by one of the libraries participating in the pilot program.[125]

The federal government was not happy when it learned that Swartz had downloaded metadata, docket sheets, and court records that accounted for roughly twenty percent of the information stored in PACER.[126] Government officials promptly suspended PACER's free-access pilot program, "pending an evaluation."[127] A few weeks later, "a Government Printing Office official . . . told librarians that 'the security of the Pacer service was compromised [and the F.B.I. was] conducting an investigation'" into the matter.[128] Nevertheless, the data that Swartz had downloaded from PACER found its way to Malamud's website, where it remains free to download.[129] The FBI eventually dropped its investigation into Swartz's mass download,[130] perhaps realizing that

---

120. Schwartz, *supra* note 110.

121. *Id.*

122. Schultze, *supra* note 105.

123. *Id.*

124. *Id.*

125. *Id.*

126. Schwartz, *supra* note 110 (reporting that Aaron Swartz, "a 22-year-old Stanford dropout and entrepreneur . . . managed to download an estimated 20 percent of the entire database: 19,856,160 pages of text").

127. *Id.*

128. *Id.*

129. *See* Bulk.Resource.Org, https://bulk.resource.org/courts.gov/ (last visited July 1, 2015).

130. Aaron Swartz, *Wanted by the FBI*, Aaron Swartz's Raw Thought (Oct. 5, 2009), http://www.aaronsw.com/weblog/fbifile.

Swartz did not break the law by downloading public records from PACER and reposting them to the web.

The recent efforts of Malamud, Schultze, Swartz, and countless other open-access advocates have reignited the debate over whether PACER should be accessible to the public free of charge. Sensing that this would not happen anytime soon, Schultze and a team of Princeton undergraduate students created a clever web application called "RECAP"—a name derived from "turning PACER around"—that users can download and run from within some web browsers when they log into PACER.[131] For every court record that a user downloads from PACER, the program saves a duplicate copy to its own database, which it then makes available to the public for free.[132] In return, RECAP alerts users when they come across court records in PACER that are available for free in RECAP's database, thereby saving the user the expense of having to download the document from PACER.[133] Since creating the program, Schultze estimates that RECAP has freed millions of records previously trapped behind PACER's pay wall.[134] As a result, startups are now able to develop and tinker with new web applications that let lawyers leverage PACER's litigation data on a large scale at little to no cost.

b. PACER Alternatives

No single company has had a more profound impact on disseminating the rule of law than West Publishing Company. Established in 1872 by two brothers, John and Horatio West,[135] the company now dominates the market for legal information. Over the last few decades, the publisher's model for distributing legal information has changed dramatically due to technological innovations. In 1973, Lexis shook up West Publishing Company's comfortable monopoly in the legal-publishing space by introducing the world's first database of electronic legal texts—LexisNexis.[136] Two years later, West introduced its own electronic case database—Westlaw—beginning the race to digitize the law.[137]

---

131. *Turning PACER Around*, RECAP L. (Aug. 14, 2009), https://www.recapthelaw.org/2009/08/14/welcome/.

132. *Id.*

133. *Id.*

134. Josh Blackman, *JoshLive: Interview with Steve Schultze About PACER, RECAP, and Access to Justice*, YOUTUBE (Apr. 3, 2013), https://www.youtube.com/watch?v=lX28yd6ks3U (statement beginning at 17:07).

135. *Company History: Historical Highlights from Across Thomson Reuters*, THOMSONREUTERS, http://thomsonreuters.com/en/about-us/company-history.html (last visited July 1, 2015).

136. Erin Carlyle, *Westlaw Rises to Legal Publishing Fame by Selling Free Information*, CITY PAGES (Apr. 29, 2009), http://www.citypages.com/2009-04-29/news/westlaw-rises-to-legal-publishing-fame-by-selling-free-information/full/.

137. *Id.*

The next major paradigm shift in legal publishing occurred with the widespread adoption of web browsers that provided lawyers with instant, remote access to a wealth of electronic legal information that previously required a trip to the courthouse. Westlaw and Lexis quickly moved their respective digital content to the web and began charging law firms considerable subscription fees for access.[138] Rather than developing sophisticated natural-language search algorithms to help lawyers more effectively identify relevant sources, Westlaw simply applied its print-based topical index, called "KeyCites," to its web-based legal content, which has since grown into a hyperlinked matrix of more than 100,000 topics and subtopics.[139] Even today, it is clear that Westlaw still views itself as a traditional publishing company, not a technology company. Rather than developing and refining machine-learning algorithms capable of automatically indexing new cases with a high degree of accuracy as they are added to its databases, Westlaw still employs an army of less efficient and more costly legal experts to manually sift through, summarize, and classify each source before making it available online.[140]

Having fattened on the fruits of their duopoly over the last couple of decades, Westlaw and Lexis's failure to innovate is hardly surprising.[141] Rather than investing in new features and competing on the basis of technology, the two companies have, in the past, simply used their deep cash reserves to acquire promising startups that might one day pose a competitive threat.[142]

---

138. *West Group History*, FUNDINGUNIVERSE, http://www.fundinguniverse.com/company-histories/west-group-history/ (last visited July 1, 2015); *About LexisNexis: Historical Milestones*, LEXISNEXIS, http://www.lexisnexis.com/en-us/about-us/about-us.page (last visited July 1, 2015).

139. Susan Nevelow Mart, *The Case for Curation: The Relevance of Digest and Citator Results in Westlaw and Lexis*, 32 LEGAL REFERENCE SERVS. QUARTERLY 13, 14, 18–19 (2013).

140. *See id.* at 18–19.

141. *See* David Hall, *Google, Westlaw, LexisNexis and Open Access: How the Demand for Free Legal Research Will Change the Legal Profession*, 26 SYRACUSE SCI. & TECH. L. REP. 53, 56 (2012) ("West and Lexis have struggled to keep up with the times and have been slow to react to customers' evolving expectations with electronic research. A large part of their reluctance to change was due to their position as a firmly entrenched duopoly.").

142. *See generally* KENDALL F. SVENGALIS, LEGAL INFORMATION BUYER'S GUIDE AND REFERENCE MANUAL 8–15 (2005) ("In the last twenty years of the 20th century, the world of legal publishing has been dramatically altered as a consequence of corporate acquisitions and mergers. While small legal publishers continue to enter the fray, many leading legal publishing houses have been acquired by major international conglomerates."). Between 1980 and 2005, Thomson Corporation had acquired over a dozen competitors in the legal-information space:

> Over the course of about twenty-five years, the Thomson Corporation, of Canada, acquired Callaghan and Company, Clark Boardman, Warren, Gorham & Lamont, Lawyers Cooperative, Bancroft-Whitney, Research Institute of America, Practitioners Publishing, Counterpoint Publishing, Gale Research,

But a new breed of legal-technology startups are emerging that offer a promising glimpse into the future of legal research. It is a future where web applications allow attorneys to not only search for and sift through legal authority, but also to learn about the tactics of their opposing counsel, the tendencies of their judge, and other behavioral aspects of litigation that tend to influence the outcome and cost of litigation ahead of time. In fact, lawyers in certain practice areas have already begun to make important strategic decisions based on trends and patterns in litigation data revealed through their use of these services.

Leading the way in the emerging renaissance in legal research is a web-startup named Lex Machina, which is Latin for "law machine."[143] Lex Machina[144] was founded in 2006 as an interdisciplinary project between the Stanford Law School and Stanford University's computer science department to bring transparency to intellectual-property litigation.[145] Today, attorneys handling bet-the-company patent litigation use Lex Machina's web-based analytics service to uncover trends and patterns in historical patent litigation to more accurately forecast costs and more effectively evaluate various case strategies.[146]

The service scrapes PACER for new cases involving intellectual property on a nightly basis.[147] Whereas Westlaw manually assigns KeyCites to the legal opinions stored within its databases,[148] Lex Machina uses a combination of human reviewers and a proprietary

> Information Access, Barclays Law Publishers, Carswell (of Canada), the West Publishing Company, Federal Publications, Inc., Findlaw, one the Internet's leading legal index and information sites, and the Harrison Company. West itself had already acquired Banks-Baldwin to go with its earlier purchases of Foundation Press, the American Law Book Company, the Edward Thompson Company, and Boston Law Book, among others.

Id.

143. Phillip J. Bond, *Guest Column: 'Law Machine' Puts Patent Files at Congress's Fingertips*, THE HILL (Dec. 23, 2013, 10:48 AM), http://thehill.com/policy/technology/193866-guest-column-law-machine-puts-patent-files-at-congresss.

144. Lex Machina was formerly referred to as the "Stanford Intellectual Property Litigation Clearinghouse." *An Empirical Study of Certain Settlement- Related Motions for Vacatur in Patent Cases*, LEX MACHINA, https://lexmachina.com/an-empirical-study-of-certain-settlement-related-motions-for-vacatur-in-patent-cases/ (last visited July 1, 2015).

145. *Stanford Law School Spins off New Venture, Offering Commercial Access to Intellectual Property Legal Data*, STAN. L. SCH. (Jan. 5, 2010), https://www.law.stanford.edu/news/pr/stanford-law-school-spins-off-new-venture-offering-commercial-access-to-intellectual-property-legal-data; Bond, *supra* note 143.

146. *See What We Do*, LEX MACHINA, https://lexmachina.com/what-we-do/ (last visited July 1, 2015).

147. Tam Harbert, *Supercharging Patent Lawyers with AI*, IEEE SPECTRUM (Oct. 30, 2013, 9:00 AM), http://spectrum.ieee.org/geek-life/profiles/supercharging-patent-lawyers-with-ai.

148. Mart, *supra* note 139, at 18–19.

algorithm that automatically parses and classifies the outcome for each case that the service extracts from PACER.[149] After extracting, processing, and scrubbing the data, Lex Machina assembles and presents aggregated case data for a particular judge, party, attorney, or law firm with analytics that allow users to quickly discern trends and patterns in the data that may affect the cost or outcome of their case.[150]

A number of other recent startups have similarly pledged to build applications that provide attorneys with "a new view on legal research."[151] Ravel Law, for example, is a search engine for legal opinions that displays the citation network for a given legal concept.[152] While Westlaw and Lexis stack search results in a column, often burying important case pages back in search results,[153] Ravel Law visually represents the most important case on a particular topic as a hub, with numerous spokes pointing to subsequent cases that have cited it.[154] The size of the hub reflects the relative number of cases that cite it.[155] The frequency with which courts cite a particular case often signals that case's influence over a given area of law or relevance to a particular legal concept.[156] Until Ravel Law's launch in 2012,[157] lawyers using Westlaw could not glean such insights into the federal Judiciary's citation networks without manually plotting each case citation by hand. Whereas manual plotting would have taken days and perhaps even weeks, Ravel Law presents its results in a matter of seconds.

Another noteworthy startup is Judicata, which is attempting to develop a Google-like search engine for finding relevant legal authority.[158] The ambitious startup will attempt to compete directly with

---

149. Harbert, *supra* note 147.

150. *See How It Works*, LEX MACHINA, https://lexmachina.com/what-we-do/how-it-works/ (last visited July 1, 2015).

151. *See* Simon Fodden, *Legal Upstarts*, SLAW (Oct. 8, 2013), http://www.slaw.ca/ 2013/10/08/legal-upstarts/.

152. Nicole Cudiamat Minnis, *The New Competition—Emerging Legal Technologies out of Silicon Valley*, NAT'L L. REV. (Feb. 16, 2015), http://www.natlawreview.com/article/new-competition-emerging-legal-technologies-out-silicon-valley-0; J.B. Ruhl, *Deep Structure—The Next Generation of Empirical Legal Studies*, LAW 2050 (Mar. 15, 2013), http://law2050.com/ 2013/03/15/deep-structure-the-next-generation-of-empirical-legal-studies/; *see also About Us*, RAVEL LAW, https://www.ravellaw.com/about (last visited July 1, 2015).

153. *See* Ruhl, *supra* note 152.

154. Robert Ambrogi, *Visual Law Services Are Worth a Thousand Words—And Big Money*, A.B.A. J. (May 1, 2014, 11:40 AM), http://www.abajournal.com/magazine/article/ visual_law_services_are_worth_a_thousand_words--and_big_money.

155. *Id.*

156. *See id.*; Ruhl, *supra* note 152.

157. *About Us*, *supra* note 152.

158. *See* Blake Masters, *Judicata: The Path of the Law*, BLAKE MASTERS (Dec. 11, 2012), http://blakemasters.com/post/37718729412/judicata-raises-2m.

Lexis and Westlaw by offering a more refined search experience that utilizes "highly specialized case law parsing and algorithmically assisted human review to turn unstructured court opinions into structured data."[159] The company's goal is to then "leverage that data to build legal research and analytics tools that are an order of magnitude better than existing offerings."[160]

The sudden explosion of technology startups focused on utilizing data to improve the delivery and consumption of legal services is remarkable for several reasons. First, for decades, the legal profession experienced very little technological disruption. While new technologies like fax, computers, electronic databases, and mobile devices enabled lawyers to become more *efficient*, they did not help them to become more *effective*. Therefore, and moreover because these developments were incremental and intermittent, they went virtually unnoticed by the profession. Christina Farr of *Venture Beat* noted that "[f]or all the talk of innovation in Silicon Valley," law has been the "one industry that has remained largely untouched by technology."[161]

But things are starting to change as data continues to play an increasingly important role in the decision-making process for lawyers and their clients. After returning from a legal-technology conference held in early 2014, Robert Ambrogi—a Massachusetts lawyer who "has been writing about legal technology for more than two decades"—wrote that he was "convinced that [he had] never before seen a time of such intense and creative innovation in legal technology."[162]

The proliferation of legal-technology startups today is also remarkable because of the sudden spike in investment activity that has made it possible. For years, investors perceived the legal-technology startup market as one of the most unattractive markets in which to invest.[163] In fact, in 2012, *TechCrunch* categorized startups based on their target market and ranked them based on the amount of capital that each category of startups had raised in recent years.[164] *TechCrunch* reported that legal startups were "in last place" in terms of new-venture fundraising, raising

---

159. *Id.*

160. *Id.*

161. Christina Farr, *Meet the Startups That Are Giving Everyone Affordable Access to Justice*, Venture Beat (Mar. 20, 2012, 10:16 AM), http://venturebeat.com/2012/03/20/legal-startups/.

162. Robert Ambrogi, *Time of Unprecedented Innovation in Legal Technology*, LawSites (Feb. 10, 2014), http://www.lawsitesblog.com/2014/02/time-unprecedented-innovation-legal-technology.html.

163. *See e.g.*, Sameer Al-Sakran, *Getting Rich by the Numbers, A CrunchBased How-To*, TechCrunch (July 21, 2012), http://techcrunch.com/2012/07/21/getting-rich-by-the-numbers-a-crunchbased-how-to/.

164. *Id.*

a mere $104 million in startup funding.[165] In stark contrast to the dismal investment figures published by *TechCrunch*, Thomson Reuters more recently reported that an "estimated $458 million . . . has been invested in legal startups in recent years . . . with an average valuation of $4.3 million" per startup.[166]

The list of investors who are now pouring money into legal-technology startups reads like a *Who's Who* of notable technology investors. Lex Machina, for example, has already raised a reported $6.8 million in funding[167] from a pool of investors that includes, among others:

- Jerry Yang, founder of Yahoo! and founding partner of AME Cloud Ventures;
- Naval Ravikant, cofounder of AngelList;
- Jeff Hammerbacher, founder and chief scientist at Cloudera;
- Joe Lonsdale, founder of Palantir Technologies and Addepar and founding partner of Formation 8;
- Radhika Shah, founder and CEO of SocialVilla;
- Eugene Zhang, founder and president of Innospring;
- Costanoa Venture Capital;
- Cue Ball Capital;
- Stanford University;
- Weiying Ding, partner at TEEC Angel Fund;
- Clint Korver, partner at Ulu Ventures;
- Lenny Mendonca, director of McKinsey & Company;
- Bill Tobin, partner at Strayer Consulting Group;
- Chris Byrne, vice president of intellectual property strategy at Samsung Innovation;
- David Chao, cofounder and general partner at DCM;
- Richard Chen, trustee of the U.C. Berkeley Foundation;
- Dan Cooperman, of counsel at DLA Piper;
- Eric Goldman, law professor at Santa Clara University School of Law;
- Mark Haynes, partner at Haynes Beffel & Wolfeld;
- Shane Johnson, managing director of PaintBrush Capital Partners;
- Brad Jones, partner at Redpoint Ventures; and
- Jean Kovacs, president of the Northern California

---

165. *Id.*

166. David Curle, *Startups: $458 Million in Legal Services R&D*, Legal Current (Feb. 11, 2014), http://legalexecutiveinsights.com/startups-438-million-in-legal-services-rd/.

167. *Lex Machina*, AngelList, https://angel.co/lex-machina (last visited July 1, 2015).

Harvard Business School Alumni Angels.[168]

Since 2012, Judicata has raised nearly $8 million in venture funding from a host of big-name investors, including:

- Peter Thiel, cofounder of PayPal, billionaire investor at Founders Fund and Clarion Capital, and former attorney;
- Keith Rabois, partner at Khosla Ventures and former attorney;
- David Lee, managing member at SV Angel; and
- David Perla, managing director at 1991 Ventures.[169]

II. BARGAINING IN THE SHADOW OF BIG DATA

This Part discusses the nature of legal "predictions" that big data will enable and attempts to define the value and limitations of such data-driven predictions. The Part also contains the crux of this Article's argument: that big data significantly enhances Coasean bargaining by lowering information costs, which traditionally formed the most easily identifiable transaction costs for purposes of Coasean analysis of legal transactions and disputes.

A. *Understanding the Lawyer–Machine Symbiosis*

To appreciate the power and limitations of software as a tool for gleaning insights from aggregated legal data, it is helpful to consider the application of data-crunching software to another domain: chess.[170] For centuries, chess was a game of human intellect—a yardstick for measuring the human mind. People used words like "genius"[171] and "prodigy"[172] to describe those who had demonstrated mastery of the

---

168. *Board, Investors & Advisors*, LEX MACHINA, https://lexmachina.com/about/board-investors/ (last visited July 1, 2015); *Lex Machina*, *supra* note 167. Note that some of the individuals in this list invested in Lex Machina on behalf of a principal while others invested in their personal capacities. This Article includes the institutional affiliations of some of the investors who fall into the latter category solely as a point of reference.

169. *Judicata*, ANGELLIST, https://angel.co/judicata (last visited July 1, 2015).

170. A portion of this Section expands upon the ideas presented in a lecture given by the cofounder of Palantir Technologies at Stanford University. During this lecture, the cofounder traced the roots of modern-day predictive analytics to early computer programs capable of beating humans at chess. *See generally* Stephen Cohen, *The Path to Palantir*, ECORNER (Feb. 20, 2013), http://ecorner.stanford.edu/authorMaterialInfo.html?mid=3052.

171. *See, e.g.*, Jay Boyar, *Convincing Story of Chess Genius*, ORLANDO SENTINEL (Sept. 17, 1995), http://articles.orlandosentinel.com/1995-09-17/entertainment/9509110220_1_bobby-fischer-josh-waitzkin-steven-zaillian.

172. *See, e.g.*, Dean Chang, *Shining Knight of Chess? One-Time Prodigy Checks out Future*, N.Y. DAILY NEWS (Sept. 24, 1995), http://www.nydailynews.com/archives/news/shining-knight-chess-one-time-prodigy-check.

game. As computers became increasingly powerful, chess players began to ask whether a computer could out-maneuver a human opponent in a chess match. In 1997, computer scientists at I.B.M. accepted the challenge by "teaching" a computer, dubbed "Deep Blue," to play chess.[173] The scientists trained Deep Blue by feeding it loads of data from previous matches played by chess grandmasters.[174] When Deep Blue faced off against chess Grandmaster Garry Kasparov later that year, the computer seemed to predict the his every move,[175] beating him in what marked the first time in history that the world's best chess player had lost to a computer under tournament conditions.[176]

As computers became increasingly powerful over the following years, the question of whether computers could outperform humans in processing information became rather mundane. Instead, people began to wonder whether the combination of humans and computers working together could outperform a super computer working alone. In 2005, Playchess.com, an online chess website, tested this theory by hosting a "freestyle chess tournament" in which players of varying skill levels formed teams that could use any resource at their disposal, including computers running chess software.[177] To the shock and amazement of the chess world, two amateur chess players from New Hampshire, using off-the-shelf computers and chess software, won the tournament, handily defeating some of the world's best chess players who were equipped with the world's most powerful chess software.[178]

So what set them apart from the best players in the world equipped with the most powerful chess software in world? As Garry Kasparov would later explain, "[t]heir skill at manipulating and 'coaching' their computers to look very deeply into positions effectively counteracted the superior chess understanding of their grandmaster opponents and the greater computational power of other participants."[179] In other words, the average players were able to leverage their computers to a much greater degree than their higher-ranking opponents because they were keenly aware of the limitations of relying on their own personal experience and intuition, yet they also recognized the human component that is so vital

---

173. Garry Kasparov, *The Chess Master and the Computer*, N.Y. Rev. Books (Feb. 11, 2010), http://www.nybooks.com/articles/archives/2010/feb/11/the-chess-master-and-the-computer/.

174. Referring to Kasparov's 1997 match against Deep Blue, IBM's website boasts that "Kasparov isn't playing a computer, he's playing the ghosts of grandmasters past." *Frequently Asked Questions: Deep Blue*, IBM, https://www.research.ibm.com/deepblue/meet/html/d.3.3a.shtml (last visited July 1, 2015).

175. *See* Kasparov, *supra* note 173.

176. *Id.*

177. *Dark Horse ZackS Wins Freestyle Chess Tournament*, Chess News (June 19, 2005), http://en.chessbase.com/post/dark-horse-zacks-wins-freestyle-che-tournament.

178. *Id.*

179. Kasparov, *supra* note 173.

to leveraging the computational strength of computers when making decisions.[180]

While this insight may have seemed novel and profound to the world of chess at the time, decades earlier Professor J. C. R. Licklider had described the working relationship between humans and computers that explains the chess amateurs' success:

> In the anticipated symbiotic partnership, men will set the goals, formulate the hypotheses, determine the criteria, and perform the evaluations. Computing machines will do the routinizable work that must be done to prepare the way for insights and decisions in technical and scientific thinking. Preliminary analyses indicate that the symbiotic partnership will perform intellectual operations much more effectively than man alone can perform them.[181]

What is remarkable about Professor Licklider's article is that, despite the dramatic changes in technology since its publication, the administrative tasks that waste valuable time and Licklider's vision for how computers might eliminate such waste remain just as prescient today. Licklider offered the following description of his typical workday in trying to solve a complex problem:

> About 85 per cent of my "thinking" time was spent getting into a position to think, to make a decision, to learn something I needed to know. Much more time went into finding or obtaining information than into digesting it. Hours went into the plotting of graphs, and other hours into instructing an assistant how to plot. When the graphs were finished, the relations were obvious at once, but the plotting had to be done in order to make them so.[182]

Most lawyers have probably experienced similar frustrations when trying to discern winning arguments from losing arguments by reading dozens of opinions over the course of several days or even weeks at their client's expense. Even more lawyers likely do not even bother to do this, opting to go with their gut instead. Consider, for example, a hypothetical attorney who is considering whether filing a motion to dismiss early in litigation would be a prudent decision. Before an informed prediction about the likely cost or result of such a motion can be made, the attorney must perform a number of mechanical tasks better left to a computer. Such tasks might include searching Westlaw for similar cases, noting the

---

180. *See* Ari Gesher, *Friction in Human-Computer Symbiosis: Kasparov on Chess*, PALANTIR BLOG (Mar. 8, 2010), http://www.palantir.com/2010/03/friction-in-human-computer-symbiosis-kasparov-on-chess/.

181. J. C. R. Licklider, *ManComputer Symbiosis*, HFE-1 IRE TRANSACTIONS ON HUMAN FACTORS IN ELECTRONICS 4, 4 (1960).

182. *Id.* at 6.

arguments made in each case, and then plotting the outcomes of each case to determine the likelihood that the presiding judge will grant their motion to dismiss, which could take hours or even days to accomplish. Only then will the attorney be in a position to make a well-informed prediction about the likelihood of success, forecast the fees and costs associated with preparing such a motion, and communicate these considerations to their client. At this point, the attorney will have already exhausted a considerable amount of resources, raising the stakes of the decision.

Unfortunately, as the above hypothetical illustrates, it currently takes far too long to search for and compile the information needed to make an informed prediction about the best course of action. Services like Lex Machina present a solution to this problem by relying on lawyers, judges, and scholars to do what they do best—coming up with hypotheses (for example, the U.S. District Court for the Eastern District of Virginia has the highest patentee "win rate")[183]—while letting computers do what they do best—testing these hypotheses against a large collection of historical data (for example, the U.S. District Court for the Eastern District of Texas has the highest patentee win rate (40.3%) while the Eastern District of Virginia's patentee "win rate" is lower (30.4%), but still higher than the U.S. District Court for the Western District of Wisconsin (24.0%)).[184] Note that, in the lawyer–computer symbiosis illustrated in this example, it is the attorney, not the computer, who must apply her domain expertise to determine which questions are important to her client's case, which data might contain answers to such questions, and which outcomes are favored.

B. *How Big (Legal) Data Changes the Game*

As mentioned in the Introduction, this Article proposes that big data differs qualitatively from the technological changes of the last few decades. Word processors, photocopiers, billing software, email, and online case databases mostly streamlined tasks that lawyers were already doing. Lawyers were sending notes and letters before email, making carbon copies before Xerox, and drafting legal documents long before WordPerfect. A few tasks gave way to complete automation, but most merely became more efficient, more replicable, and used less space.

Despite this electronic streamlining, the approach to lawyering was much the same. Legal rules dominated the approach to practice. Legal advice would center on the rules—case law, statutes, and regulations—and their mandated outcomes. Legal analysis meant applying the law to

---

183. Robert Angle, *Eastern District of Virginia—A Top Venue for Patentee Plaintiffs (?)*, VA. IP L. (May 2010), http://virginiaiplaw.com/2010/05/eastern-district-of-virginia-a-top-venue-for-patentee-plaintiffs/.

184. *See id.*

the facts and making comparisons, distinctions, analogies, and normative arguments. Legal predictions were assessments of the probability that a client's situation—his set of facts—would trigger a particular legal rule, thereby compelling a given outcome. Legal advice instructed clients how to work within or around the rules. Where the rules conferred broad discretion on judges or regulators, lawyers had to cope with uncertainty.

Most cases settle before trial, and most transactions do not result in litigation.[185] Even so, negotiations and bargaining, whether over contracts, plea bargains, or lawsuit settlements, have traditionally taken place in the shadow of the trial, as most cases are resolved in the lead up to trial based on the parties' and lawyers' expected outcome at trial, discounted by the money that they expect to save by not having to proceed to trial.[186] The technologies that lawyers historically embrace still function within this paradigm. Online case research, conference calls, drafting documents on laptops, and scheduling with smartphones all still take place in the shadow of the trial.

Big data moves lawyers away from the shadow of the trial and into the shadow of data. Software designed to surface trends and patterns in legal data as described above provide attorneys with powerful tools for gleaning relevant, practical insights into dynamics that affect litigation beyond the black-letter law. A judge's tendency to grant or deny certain motions may tell the lawyer more about his next decision than the statutory verbiage governing the circumstances.[187] Similarly, which judges or jurisdictions, if any, moved dockets along quickly, micromanaged settlements or discovery, or loved or eschewed special masters—over the last year or last decade—is data that lawyers can now obtain and consider. Big data yields clear behavioral profiles, charted across time, of opposing parties, opposing counsel, judges in an alternate forum, and so on. Discretionary rulings shed their uncertainty and become a matter of pattern matching and predictions. Big data changes

---

185. George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. LEGAL STUD. 1, 2 (1984).

186. *See* HARRY KALVEN JR. ET AL., THE AMERICAN JURY 31–32 (1966); Thomas W. Church Jr., *In Defense of "Bargain Justice,"* 13 L. & SOC'Y REV. 509, 512–14, 523 (1979); Robert Cooter et al., *Bargaining in the Shadow of the Law: A Testable Model of Strategic Behavior*, 11 J. LEGAL STUD. 225, 225 (1982); Frank H. Easterbrook, *Criminal Procedure as a Market System*, 12 J. LEGAL STUD. 289, 311–17 (1983); Frank H. Easterbrook, *Plea Bargaining as Compromise*, 101 YALE L.J. 1969, 1975 (1992); William M. Landes, *An Economic Analysis of the Courts*, 14 J.L. & ECON. 61, 66, 69 (1971); Mnookin & Kornhauser, *supra* note 16, at 968; Priest & Klein, *supra* note 185, at 2; Edward A. Ruttenberg, *Plea Bargaining Analytically: The Nash Solution to the Landes Model*, 7 AM. J. CRIM. L. 323, 353 (1979); Robert E. Scott & William J. Stuntz, *Plea Bargaining as Contract*, 101 YALE L.J. 1909, 1909 (1992); Steven Shavell, *Suit, Settlement, and Trial: A Theoretical Analysis Under Alternative Methods for the Allocation of Legal Costs*, 11 J. LEGAL STUD. 55, 56 (1982).

187. *See* Charles J. Snyder, *Moneyball Lawyering*, 65 ARK. L. REV. 837, 862 (2012).

the traditional approach to lawyering because it adds a large dimension besides the legal rules to legal decision-making. Of course, the computer will not tell the lawyer what to do, but it completely changes the context for making strategic decisions.

When lawyers refer to relying on their "instinct" or "experience," the legal profession assumes that they are not talking about the legal rules. These lawyers are not claiming that they trust hunches about statutes or case precedents that they could simply look up, but rather their hunches about the leanings of a particular judge or jury, the resolve of the opposing party or lawyer, or the salesmanship of a witness. Such intuitions are experience-based inferences about the tendencies of certain individuals to behave in certain ways drawn from the lawyer's personal observations or discussions with her colleagues. In other words, in addition to looking up legal rules to ascertain the mandated disposition of an issue, lawyers also engage in unscientific profiling of the other characters involved in a matter. Therein lies the main value, both to the client and to prospective lateral-hire employers, of legal experience—the rather intangible knowledge about how things work and how key players in a legal matter behave.

Up to now, technological advances in law practice had almost nothing to do with this component of lawyering—the lawyer's experience-based inferences, hunches, or predictions. Instead, rolodexes moved to databases, bookkeeping moved to spreadsheets, meetings moved to conference calls, and so on. Certain manual tasks gave way to automation. The software enabling lawyers to make decisions based on data differs from previous technological advances in that it directly relates to experience-based inferences, except now, a lawyer can glean insights in minutes when this skill previously would have taken years of observations and personal interactions to acquire.

C. *Coasean Bargaining in the Shadow of Big Data*

A feature of big data largely ignored in the legal and social science literature up to now is its potential to affect bargaining and negotiations by lowering information costs and information asymmetries. Nearly everything written about big (legal) data has been in either marketing material[188] or policy literature discussing privacy concerns.[189]

---

188. *See, e.g.*, Owen Byrd, *Moneyball for Lawyers: How Data and Analytics Are Transforming the Practice of Law*, BOTTOM LINE (Apr. 2013); Sharon Driscoll, *A Positive Disruption: The Transformation of Law Through Technology*, STAN. LAWYER (July 4, 2013), http://stanfordlawyer.law.stanford.edu/2013/06/a-positive-disruption/ (discussing different legal technology companies started by Stanford graduates); Masters, *supra* note 158 (promoting his research company, Judicata).

189. *See, e.g.*, Julie E. Cohen, *What Privacy Is For*, 126 HARV. L. REV. 1904, 1918–19, 1927 (2013) (discussing big data as the latest threat to personal privacy and individual liberty);

Analyzed under a Coasean point of view,[190] lower transaction costs increase the chance of bargaining while "mak[ing] legal rules in the surrounding environment less important for the eventual outcome."[191] Information costs and asymmetries are primary transaction costs for all forms of lawyering outside the courtroom[192]—contract negotiation and drafting, settlement talks, plea-bargaining, and the full range of transactional work.[193] Lower transaction costs make negotiating under, or around, the legal rules easier.[194] The Coase Theorem suggests[195] that "bargains are more likely to occur when transaction costs are low,"[196] and

---

Woodrow Hartzog & Frederic Stutzman, *The Case for Online Obscurity*, 101 CALIF. L. REV. 1, 1 (2013) ("[O]bscurity is a critical component of online privacy."); Lee Tien, *Privacy, Technology and Data Mining*, 30 OHIO N.U. L. REV. 389, 391 (2004) (discussing "whether the Fourth Amendment constrains government data mining in any meaningful way"); Peter A. Winn, *Judicial Information Management in an Electronic Age: Old Standards, New Challenges*, 3 FED. CTS. L. REV. 135, 136, 141 (2009) (discussing "the changes which have come about with the introduction of electronic information systems and the death of practical obscurity").

190. *See* R. H. Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1 (1960). R.H. Coase's groundbreaking article, *The Problem of Social Cost*, helped launch the law and economics movement among legal scholars. Citations to Coase's work have become ubiquitous in academic scholarship. *See generally* Daniel Farber, *Parody Lost/Pragmatism Regained: The Ironic History of the Coase Theorem*, 83 VA. L. REV. 397 (1997) (providing an excellent overview of literature based on the Coase Theorem).

191. Dru Stevenson, *Costs of Codification*, 2014 U. ILL. L. REV. 1129, 1132 & nn.9–10 [hereinafter Stevenson, *Costs of Codification*]. "The Coase Theorem posits roughly that legal rules or rights matter least where parties have the most opportunity to negotiate; conversely, rules and rights matter more when parties have less opportunity to bargain around the laws." *Id.* at 1132 n.10. This is Professor Stevenson's paraphrasing; Coase's argument was not simplified down into one sentence in his original article. *Id.* Coase attributed the "Coase Theorem" moniker to economist Joseph Stigler. *See* R. H. COASE, THE FIRM, THE MARKET, AND THE LAW 157 (1988) [hereinafter COASE, THE FIRM].

192. *See* Eugene Kontorovich, *The Constitution in Two Dimensions: A Transaction Cost Analysis of Constitutional Remedies*, 91 VA. L. REV. 1135, 1150 n.29 (2005) ("Transaction costs, broadly defined, sometimes purchase real benefits. Some types of transaction costs—such as the costs of negotiating contracts, which involve obtaining information about the value of the thing being contracted for—allow prospective purchasers to evaluate whether the transaction should proceed.").

193. *See* George M. Cohen, *The Fault Lines in Contract Damages*, 80 VA. L. REV. 1225, 1242 (1994) ("Transaction costs involved in making contracts include costs of collecting information, negotiating, and drafting.").

194. Coase, *supra* note 190, at 15 ("It is always possible to modify by transactions on the market the initial legal delimitation of rights. And, of course, if such market transactions are costless, such a rearrangement of rights will always take place if it would lead to an increase in the value of production.").

195. "Coase's article is widely considered to be the most-cited article of all time in legal scholarship." Stevenson, *Costs of Codification*, *supra* note 191, at 1150 n.127.

196. *Id.* at 1150. But transaction costs do not actually have to be low. A purely rational actor will invest $99.99 to get $100 from the government. *See* Jonathan R. Macey, *Transactional Costs and the Normative Elements of the Public Choice Model: An Application to Constitutional Theory*, 74 VA. L. REV. 471, 478–79, 511 (1988). Yet this transaction is not necessarily efficient.

numerous scholars have identified information costs as a primary category of transaction costs overall.[197] Big data lowers information costs and information asymmetries between parties; bargains are therefore more likely to occur as they become easier to initiate and consummate.

For example, in the context of litigation and settlement negotiations, parties are far more likely to reach an agreement to settle before trial—thereby saving both sides the costs and anxiety of litigation—if each side has better information about their opponent.[198] This would include information about the opposing party and their litigation history, as well as opposing counsel and his track record and patterns of decision-making, all of which are now available through big data. Particularly important would be information about previous settlement agreements—whether involving the opposing party, their counsel, or a related nexus or cause of litigation—and this information is unavailable through the conventional Westlaw and Lexis services. Professor Ben Depoorter has recently argued that information about previous settlements is supremely important for those trying to settle a new case before trial,[199] and the data-analytic firms provide invaluable information in this regard.[200] Even post-verdict settlement negotiations, which seem to occur more frequently than previously thought,[201] depend on information and are, in turn, a valuable source of information for future litigants.[202] Settlement negotiations are a classic case of Coasean bargaining,[203] and big data facilitates such negotiations by providing both parties with better information. Not only do the parties have less uncertainty to interfere with agreements, but also

---

Rather it represents a "deadweight cost" since that $99.99 could have been spent to generate wealth and instead was spent to transfer wealth. *Id.* at 478–79.

197. *See, e.g.*, Daniel B. Kelly, *Strategic Spillovers*, 111 COLUM. L. REV. 1641, 1659 n.56 (2011) ("'Coasean bargaining'—frictionless negotiation with zero transaction costs—assumes not only the absence of bargaining costs but also perfect information."); Robert C. Ellickson, *The Case for Coase and Against "Coaseanism,"* 99 YALE L.J. 611, 619–20 (1989) (describing information costs as fatal to Coasean bargaining even where other transaction costs are minimal); Cynthia Estlund, *Just The Facts: The Case for Workplace Transparency*, 63 STAN. L. REV. 351, 369–70 (2011); Richard H. McAdams, *Relative Preferences*, 102 YALE L.J. 1, 65–67 (1992) (discussing the role of information as a transaction cost that impedes Coasean Bargaining).

198. Bruce L. Hay, *Effort, Information, Settlement, Trial*, 24 J. LEGAL STUD. 29, 42–43, 62 (1995); *see also* Ben Depoorter, *The Upside of Losing*, 113 COLUM. L. REV. 817, 824 & n.30 (2013); Kathryn E. Spier, *The Dynamics of Pretrial Negotiation*, 59 REV. ECON. STUD. 93, 102–03 (1992).

199. Ben Depoorter, *Law in the Shadow of Bargaining: The Feedback Effect of Civil Settlements*, 95 CORNELL L. REV. 957, 974 (2010).

200. *See, e.g.*, Ambrogi, *supra* note 154.

201. *See* Wendy E. Wagner, *Administrative Law, Filter Failure, and Information Capture*, 59 DUKE L.J. 1321, 1370–71 (2010).

202. *See id.*

203. *See* Kelly, *supra* note 197, at 1695, 1720.

their positions are likely to move closer together as each side has a more realistic picture of the case and its likely outcome, leaving less disagreement to keep them apart.[204] Lawyering with big data means that cases are more likely to settle overall and that parties should reach agreements more quickly.[205]

Of course, an increase in settlement rates due to Coasean bargaining can present downsides as well. To the extent that settlement lowers the expected total cost of litigation, marginal plaintiffs (especially those with smaller claims) are more likely to bring suit—settlements greatly decrease the number of trials but may also increase the number of claims brought. Thus, the Coasean bargaining benefits of big data, in the context of litigation settlements, will have a partial offset as marginal plaintiffs are more likely than before to initiate new lawsuits.

Litigation is not the only arena for Coasean bargaining. Property transactions also depend on low bargaining costs, and information costs have historically been prevalent transaction costs for agreements to transfer real estate[206] or intellectual property.[207] Coasean bargaining occurs regularly between city planners and developers,[208] and the information now available through big data can significantly facilitate such bargaining—again by reducing information costs for both sides. Energy trading is a growing area of transactional work, and information costs have traditionally been a barrier to Coasean bargaining in this

---

204. *See* Steven Shavell, *Any Frequency of Plaintiff Victory at Trial Is Possible*, 25 J. LEGAL STUD. 493, 495 (1996) (explaining that cases are less likely to settle and more likely to go to trial when the parties disagree on the likely outcome).

205. *See* David G. Anderson, *Effective Construction Claim Resolution: Understanding DCAA*, 43 PUB. CONT. L.J. 165, 196 (2014) (discussing "the information needed to settle a claim quickly without discovery or trial"); Barry R. Furrow, *The Patient Injury Epidemic: Medical Malpractice Litigation as a Curative Tool*, 4 DREXEL L. REV. 41, 77 (2011) ("Regulators can act more quickly and effectively with full information on the number of lawsuits, the settlements, and information revealed in pretrial discovery."); Charles Silver, *"We're Scared to Death": Class Certification and Blackmail*, 78 N.Y.U. L. REV. 1357, 1378 (2003) ("Settlement values can also change quickly in response to new information.").

206. *See* Joseph Blocher, *Selling State Borders*, 162 U. PA. L. REV. 241, 254 (2014) (applying Coasean bargaining principles to property transactions between states); Abraham Bell & Gideon Parchomovsky, *The Case for Imperfect Enforcement of Property Rights*, 160 U. PA. L. REV. 1927, 1945–46 (2012) (discussing signage and fencing as signaling devices in the context of Coasean bargaining between potential buyers and sellers of real estate).

207. *See* Christopher Buccafusco & Christopher Sprigman, *Valuing Intellectual Property: An Experiment*, 96 CORNELL L. REV. 1, 17 (2010) (discussing Coasean bargaining and intellectual property transactions); Ronald J. Gilson et al., *Contracting for Innovation: Vertical Disintegration and Interfirm Collaboration*, 109 COLUM. L. REV. 431, 456 (2009) (applying Coasean bargaining principles to "contracting for innovation").

208. *See* David Schleicher, *City Unplanning*, 122 YALE L.J. 1670, 1682 (2012) (discussing Coasean bargaining between city planners and developers).

setting as well;[209] big data should lower these costs and streamline such transactions. Any transaction complex enough to require lawyers to consummate the deal becomes easier in the shadow of big data—valuations become more accurate and harmonized between parties, and prospective buyers and sellers can retain the benefits of arms-length transactions while gaining the assurance of information that previously came only through familiarity.

A more subtle effect of big data on lawyering, from a Coasean perspective, is that legal rules or procedures governing the situation have less import—a corollary of the Coase Theorem.[210] As big data facilitates bargaining by lowering information costs and asymmetries, a side effect is the dilution of various legal rules surrounding the subject of negotiation. Transaction costs—the things that prevent automatic agreement among parties—"are what animate the legal rules, or give them their verve."[211] The Coase Theorem implies that ex ante legal rules and entitlements are less significant when parties can easily transact around them.[212]

In the real world, Coase emphasized that transaction costs vary greatly but are never zero.[213] The import of the relevant legal rules or rights would therefore correlate with the transaction costs present in a given situation.[214] Where transaction costs are low, bargaining becomes easy, and the legal rules exert less control over the eventual allocation of resources. As big data provides parties with significantly more information and reduces uncertainty about other parties, trends in the courts, and even external transaction costs (such as the length of time

---

209. Troy A. Rule, *Property Rights and Modern Energy*, 20 Geo. Mason L. Rev. 803, 832–33 (2013) (discussing Coasean bargaining and property rights in the energy arena, and mentioning information costs as a traditional obstacle).

210. *See* Dru Stevenson, *Jury Selection and the Coase Theorem*, 97 Iowa L. Rev. 1645, 1655–58 (2012) [hereinafter Stevenson, *Jury Selection*]; Russell Korobkin, *The Status Quo Bias and Contract Default Rules*, 83 Cornell L. Rev. 608, 623 (1998) ("Economic analysis thus suggests that default rules matter in only a subset of the total number of contractual relationships. If transaction costs are low and information is distributed symmetrically between the parties, they will negotiate efficient contracts regardless of the content of default rules.").

211. Stevenson, *Costs of Codification*, *supra* note 191, at 1151; *see* Thomas W. Merrill & Henry E. Smith, *Making Coasean Property More Coasean*, 54 J.L. & Econ. S77, S78 (2011); *see also* Coase, *supra* note 190, at 15–19; Coase, The Firm, *supra* note 191, at 178 ("The same approach which, with zero transaction costs, demonstrates that the allocation of resources remains the same whatever the legal position, also shows that, with positive transaction costs, the law plays a crucial role in determining how resources are used.").

212. Stevenson, *Costs of Codification*, *supra* note 192, at 1151. *See generally* Harold Demsetz, *When Does the Rule of Liability Matter?*, 1 J. Legal Stud. 13 (1972).

213. *See* Coase, The Firm, *supra* note 191, at 174–75; Ellickson, *supra* note 197, at 612.

214. *See* Merrill & Smith, *supra* note 211, at 92–99; Stevenson, *Jury Selection*, *supra* note 210, at 1654–58.

typically involved in enforcing an agreement in the courts), a primary hurdle to agreements disappears.[215]

In sum, an important implication of lawyering in the shadow of data is that the newly available information should facilitate more bargaining, agreements, settlements, and other legal transactions. As bargaining or consummating deals is an important part of what lawyers do, big data should prove incredibly useful for practitioners who strike the right balance between intuition, personal experience, and behavioral insights gleaned from aggregated legal data. At the same time, a side effect of the reduced information costs is that it is also easier to bargain around certain legal rules and assignments of rights, which could have important policy implications beyond the scope of this Article.[216] Surprisingly, the Coasean implications of big data have so far received no identified attention in the legal academic literature, the business academic literature, or the social science literature.

D. *Ambulance Chasing with Big Data*

A recent article demonstrated how mapping data from social networking sites precisely traced the impact of an East Coast hurricane on individual neighborhoods.[217] The Center for Disease Control can monitor Google search trends to track the spread of seasonal flu outbreaks.[218] These are well-known instances where researchers knew

---

215. *See* Joseph Blocher, *Institutions in the Marketplace of Ideas*, 57 DUKE L.J. 821, 839 (2008) ("Coase . . . divided transaction costs into four categories: search, information, negotiation, and enforcement. Contrary to the neoclassical model's assumptions of perfect information and costless exchange, Coase recognized that these costs distort the market, making the mathematical predictions of neoclassical models largely inapplicable in the real world. Every transaction cost, he realized, is a small market failure." (footnotes omitted)).

216. *See, e.g.*, Barry E. Adler & Ian Ayres, *A Dilution Mechanism for Valuing Corporations in Bankruptcy*, 111 YALE L.J. 83, 148 (2001) ("Where markets are thick, information full, and transaction costs low, either markets or structured negotiation backed by judicial determination will allow a financially distressed firm to reorganize easily in a manner consistent with absolute priority."). For a similar application of transaction cost analysis to policy making outside the arena of big data, see Guy Halfteck, *Legislative Threats*, 61 STAN. L. REV. 629, 701 (2008) ("Group organization increases the likelihood of group-wide compliance and renders legislative threats more effective as catalysts of reform. Moreover, the tendency towards organization reduces the transaction costs of regulatory bargaining, consequently enabling legislators and groups to share information, negotiate, and design superior regulatory measures.").

217. Taylor Shelton et al., *Mapping the Data Shadows of Hurricane Sandy: Uncovering the Sociospatial Dimensions of 'Big Data,'* 52 GEOFORUM 167, 178 (2014), *available at* http://dx.doi.org/10.1016/j.geoforum.2014.01.006.

218. Mohammad Tufail Malik et al., *"Google Flu Trends" and Emergency Department Triage Data Predicted the 2009 Pandemic H1N1 Waves in Manitoba*, 102 CANADIAN J. PUB. HEALTH 294, 296 (2011), *available at* http://go.galegroup.com/ps/i.do?id=GALE%7CA264481478&v=2.1&u=gain40375&it=r&p=HRCA&sw=w&authCount=1; Geoff Watts, *Google Watches over Flu*, 338 BRIT. MED. J. 74, 74 (2009), *available at* http://www.jstor.org/stable/20511741?orig

what they were looking for. Big data, however, can also show analysts what to look for in the first place, as patterns and correlations emerge that nobody would have predicted. As an example of the more conventional use of big data, plaintiffs' lawyers who have heard about a major chemical spill[219] or product recall announcement can use data culled from social networking sites or search engines to find as many of the victims—potential clients—or witnesses as possible. This mimics what lawyers in previous generations would have done with newspaper notices and short television ads.

A more novel approach, however, would be to use big data to identify mass torts (or even serial tortfeasors, such as a negligent contractor or architect) that are yet unknown.[220] For example, one shipment of products to a particular store may have been a defective batch, resulting in localized or clustered injuries that are unlikely to attract media attention, or a particular scammer may have a spate of victims within a short period. In other words, big data could help identify tortfeasors as well as victims. Lawyers themselves could be the subjects of such searches—big data may eventually help state disciplinary authorities uncover unauthorized practice of law by non-lawyers or foreign lawyers (which often goes undetected now), legal malpractice, and other forms of misconduct. The availability of rich data can yield metrics to measure lawyer competence more objectively than society does at present.

On the legal defense side, it is probably unnecessary to suggest that lawyers could utilize big data to identify tortfeasors or victims, as the

---

in=JSTOR-pdf; Vanja M. Dukić et al., *Tracking Flu Epidemics Using Google Flu Trends and Particle Learning* 3, 8–10 (Nov. 25, 2009) (unpublished manuscript), *available at* http://ssrn.com/abstract=1513705; Philip M. Polgreen et al., *Using Internet Searches for Influenza Surveillance*, 47 CLINICAL INFECTIOUS DISEASES 1443, 1443, 1446 (2008), *available at* http://cid.oxfordjournals.org/content/47/11/1443.full.pdf ("With use of the frequency of searches, our models predicted an increase in cultures positive for influenza 1–3 weeks in advance of when they occurred (P< .001), and similar models predicted an increase in mortality attributable to pneumonia and influenza up to 5 weeks in advance (P< .001)."). *But see* Declan Butler, *When Google Got Flu Wrong: US Outbreak Foxes a Leading Web-based Method for Tracking Seasonal Flu*, 494 NATURE 155, 155–56 (2013), *available at* http://www.nature.com/news/when-google-got-flu-wrong-1.12413 (attributing Google's overestimates to excessive media coverage of seasonal flu outbreaks, which in turn triggered disproportionate internet searches about the topic); Nick Bilton, *Disruptions: Data Without Context Tells a Misleading Story*, N.Y. TIMES (Feb. 24, 2013, 11:00 AM), http://bits.blogs.nytimes.com/2013/02/24/disruptions-google-flu-trends-shows-problems-of-big-data-without-context/.

219. Sara Gosman, *Reflecting Risk: Chemical Disclosure and Hydraulic Fracturing*, 48 GA. L. REV. 83, 102–04 (2013) (discussing how greater information disclosures would enable more Coasean bargaining).

220. *See* Sharad Goel et al., *Predicting Consumer Behavior with Web Search*, 107 PROC. NAT'L. ACAD. SCI. 17486, 17488–89 (2010), *available at* http://www.pnas.org/cgi/doi/10.1073/pnas.1005962107 (discussing the use of big data to predict consumer behavior).

FLORIDA LAW REVIEW [Vol. 67

insurers and manufacturers that retain defense counsel are generally years ahead of the legal profession in the use of big data and analytics. Presumably, many insurers and large companies are already using big data to predict the location and scope of upcoming liabilities, just as police use it to predict the neighborhoods that need the most patrols. Lawyers could add to this analysis by using data analytics to predict what incidents will turn into class actions rather than individual lawsuits, and how much time defendants probably have before they face actual liability.

The use of big data by potential defendants (insurers and tortfeasors) leads to another likely development—the impact of data availability on liability itself.[221] Companies now have more information than before about their past, current, and future customers, making misuses of their product, as well as injuries from defects, much more foreseeable. This enhanced ability to foresee consumer injuries directly affects the duty of care, the duty to warn, the duty to recall products, and so on. It is also easier to trace lines of causation, both before and after an accident. In addition, online purchasing directly from manufacturers already eliminates many of the jointly liable co-defendants of the past, that is, retailers and distributers.

III. NORMATIVE CONSIDERATIONS

So far, this Article has explored how summarizing trends and patterns across historical data allows lawyers to make better predictions about various measurable aspects of litigation. The ability to predict litigation outcomes in particular rests on the assumption that such outcomes are not entirely random. It is hardly controversial to suggest that litigation outcomes are not entirely random given the important role that precedent plays in guiding judicial decision-making. Historical litigation data lends itself to predictive modeling because judges, particularly district court judges who cannot make new law, are generally bound by the decisions of their appellate-level predecessors—stare decisis.[222]

A. *The Predictive Value of Precedent*

The federal Judiciary's preference for predictable decisions serves several salutary interests that are vital to the preservation of a healthy legal system. First, basic fairness dictates that judges reach similar results

---

221. Bryant Walker Smith, *Proximity-Driven Liability*, 102 GEO. L.J. 1777, 1796–97, 1812 (2014).

222. *See* Stewart E. Sterk & Kimberly J. Brunelle, *Zoning Finality: Reconceptualizing Res Judicata Doctrine in Land Use Cases*, 63 FLA. L. REV. 1139, 1147 & n.27 (2011) (noting that stare decisis "often leads courts to abide by past decisions . . . [but] courts should, at least sometimes, be free to overrule past precedent.").

in similar cases.[223] Predictable decisions "foster[] reliance on judicial decisions, and contribute[] to the actual and perceived integrity of the judicial process."[224] Moreover, a judge who delivers rulings that are consistent with her previous rulings reduces legal uncertainty, which, in turn, allows individuals and organizations to plan their affairs with little fear of unanticipated legal consequences.[225] The absence of predictable outcomes, on the other hand, stifles personal growth and economic progress.[226] For these reasons and many more, courts have a vested interest in deciding issues in a manner that is consistent with precedent. This, in turn, enhances the predictive value of historical trends and patterns gleaned from litigation data analytics.[227]

The fact that laws change over time does not strip historical litigation data of its predictive value. While some laws do indeed change, the change tends to be incremental, thereby reducing the number of outlier decisions that defy precedent.

B. *Self-Fulfilling Prophecies and Self-Defeating Predictions*

Even forward-looking forecasts based on highly predictive historical data, however, can and do fail for a variety of reasons worth exploring. Some predictions can fail due to confusion from signals and noise in the data set;[228] others fail because of misinterpreted lines of causation; still others fail because of unique, unpredictable "Black Swan" events;[229] and some predictions become self-cancelling or self-fulfilling as richer information becomes available to a wider range of actors. Noise–signal problems are not unique to law and are an ever-present hazard for any superficial or amateur attempt at data analysis.[230]

---

223. *See* James B. Beam Distilling Co. v. Georgia, 501 U.S. 529, 537 (1991).

224. *See* Payne v. Tennessee, 501 U.S. 808, 827 (1991); *see also* Burnet v. Coronado Oil & Gas Co., 285 U.S. 393, 406 (1932) (Brandeis, J., dissenting) ("In most matters it is more important that the applicable rule of law be settled than it be settled right.").

225. *See* Teague v. Lane, 489 U.S. 288, 332 (1989); *see also* Randy Beck, *Transtemporal Separation of Powers in the Law of Precedent*, 87 Notre Dame L. Rev. 1405, 1408, 1428 (2012).

226. *See* Itel Containers Int'l Corp. v. Huddleston, 507 U.S. 60, 79 (1993) (Scalia, J., concurring in part and in the judgment) ("[T]he principal purposes of *stare decisis* . . . are to protect reliance interests and to foster stability in the law."); *see also* Henry M. Hart Jr. & Albert M. Sacks, The Legal Process: Basic Problems in the Making and Application of Law 568 (William N. Eskridge Jr. & Phillip P. Frickey eds., 1994).

227. *See* Andrew Stranieri & John Zeleznikow, *Knowledge Discovery from Legal Databases—Using Neural Networks and Data Mining to Build Legal Decision Support Systems*, *in* Information Technology & Lawyers: Advanced Technology in the Legal Domain from Challenges to Daily Routine 81, 87–88 (Arno R. Lodder & Anja Oskamp eds., 2006).

228. *See generally* Nate Silver, The Signal and the Noise: Why So Many Predictions Fail—But Some Don't (2012).

229. *See generally* Nassim Nicholas Taleb, The Black Swan: The Impact of the Highly Improbable (2007) [hereinafter Taleb, The Black Swan].

230. *See* Silver, *supra* note 228, at 1, 16–17.

Particularly problematic for law, however, are the self-fulfilling and self-nullifying tendencies of certain predictions. For example, there are instances where statistical predictions related to law or enforcement become self-fulfilling or reinforcing, as when police use statistics to identify high-crime neighborhoods, shift resources to those neighborhoods, and then find that the arrest numbers, and hence the "crime rate," increase or get worse.[231] This is a type of feedback effect— an informational input for decision makers skews the decisions and resultant outcomes toward producing more of the same information.

In contrast, other cases illustrate how legal data becomes self-negating, as when a court with a record of favoring plaintiffs thereby attracts plaintiffs with increasingly meritless cases, so eventually the number of cases reverts to the mean.[232] This is a type of rebound effect, where a decision maker's informational input affects the decisions and resultant outcomes in such a way as to make the information an inaccurate predictor. In other words, a disparity with strategic significance for parties, once widely known, will tend to attract marginal parties for the favored side and repel marginal parties for the disfavored side, so that outcomes eventually even out over time. Statistician-turned-columnist Nate Silver points to drivers' use of GPS traffic information to find the least congested route across Manhattan as one example of self-defeating predictions; when too many drivers follow the data about the least-congested thoroughfare, that route quickly becomes the most congested.[233]

Assuming that predictions become easier and more accurate with the advent of accessible data, a crucial question for the legal system is which predictions carry the risk of becoming self-fulfilling and which could become self-defeating. The answer seems to depend primarily on two variables: information asymmetries and binary strategic incentives. Returning to Nate Silver's work, apparently the primary factor determining which way a prediction will tip is whether the individual decision maker seeks "bandwagon benefits" or, in contrast, hopes to gain an insider advantage.[234] Politicians hope to gain bandwagon benefits in the form of additional votes by courting public endorsements from

---

231. BERNARD E. HARCOURT, AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE 156 (2007).

232. The "judicial-hellhole" phenomenon—where a streak of plaintiff wins creates a type of litigation gold rush—seems to reflect a bandwagon benefit, at least in the short term, but becomes self-nullifying over time due to the self-selection or screening effects of marginal plaintiffs and defendants. *See generally Judicial Hellholes: About*, AM. TORT REFORM FOUND, http://www.judicialhellholes.org/about/ (last visited July 1, 2015) (discussing the background of these "judicial hellholes").

233. SILVER, *supra* note 228, at 219.

234. *See id.* at 260, 341–42.

influential celebrities.[235] One study showed that, in 2008, President Barack Obama gained as many as one million additional votes from the "Oprah Effect" alone.[236] Savvy politicians know that elections are won and lost based on crowds lurching in their direction; the most effective politicians know how to garner support by jumping on a bandwagon that has momentum.

Infected patients—Silver's second example[237]—are better off when there is a higher chance of a doctor's visit yielding a quick diagnosis and readily available treatment, which is more likely when the disease is a case that the physician is seeing frequently, as in a publicized outbreak or minor epidemic.[238] Drivers, in contrast, have the exact opposite utility function—they want the road less traveled, or at least the one less congested at that moment.[239] For litigants, the same dichotomy is also supremely important in determining which way legal predictions will tip. These examples highlight how strategic incentives—especially in two-sided competitions or instances with binary choices—can turn informational richness into a self-cancelling or self-fulfilling phenomenon.

Yet information asymmetries also play a critical role in deciding the reliability of predictions, especially in determining whether a prediction will trend at all toward self-fulfillment or self-defeat or whether the original prediction will prove accurate. Bandwagon benefits tend to occur where information is widely distributed among decision makers in the same "ecosystem" (for example, traffic patterns and celebrity endorsements). Private information, on the other hand, leads to information asymmetry, which changes the dynamic by eliminating crowd responses and allowing one party to exploit the prediction to the disadvantage of another. With crowd-response examples, public information is desirable for those seeking bandwagon benefits, as bandwagon effects thrive on information feedback loops. When incentives are more competitive, rather than collaborative, public information eliminates many chances for opportunism by the parties, as everyone is on equal footing, can glean from the same predictions, and can predict the reactions of others to the original predictions.

Litigation is inherently competitive, and access to PACER data, at least on the scale required for super crunching, is prohibitively expensive

---

235. *Id.* at 47–49.

236. *See generally* Craig Garthwaite & Timothy J. Moore, *Can Celebrity Endorsements Affect Political Outcomes? Evidence from the 2008 U.S. Democratic Presidential Primary*, 29 J.L. Eco. & Org. 355 (2012).

237. *See* Silver, *supra* note 228, at 204–32.

238. *See id.* at 218–19.

239. *See id.* at 219.

for many litigants.[240] Predictions based on big (legal) data can become self-fulfilling or self-nullifying depending on the degree of information asymmetry present in a given case and the extent to which the parties' incentives are a zero-sum game. If all the parties know, for instance, that a particular judge or jurisdiction is more favorable to plaintiffs, the tendency will be for marginal defendants to avoid the forum if at all possible and for marginal plaintiffs to seek it. This, in turn, will eventually lead to fewer plaintiff-friendly outcomes as the number of frivolous cases rises. Thus, widely available information—low informational asymmetry—coupled with an incentive to beat the crowds will normally yield self-defeating predictions as rich data becomes increasingly accessible to the masses.

On the other hand, a type of bandwagon effect for lawyers can occur where a jurisdiction has a large body of precedents on an issue, as this benefits all parties by reducing uncertainty and fostering earlier settlements. In that case, the information becomes self-reinforcing. The widely available information in this scenario, combined with bandwagon benefits, yields self-fulfilling prophecies.

In litigation, as in military conflict, business, politics, and just about any other competitive facet of life, success often depends on access to information. Better intelligence leads to more accurate predictions and smarter decisions. In the context of litigation, this might include predictions about the strategic advantages and disadvantages of filing a lawsuit in a particular forum, which arguments will appeal to a particular judge, the type of tactics to anticipate from the opposing counsel, the cost of litigating a case to trial, and the list goes on. Recognizing that relevant information is a valuable resource in the adversarial context of a lawsuit, law firms and their clients are willing to pay sizeable sums for access to the vast trove of information stored in Westlaw and Lexis. Some have gone so far as to accuse wealthier litigants who are able to afford access to unpublished opinions compiled by Westlaw or Lexis as gaining an unfair advantage over their adversaries without such access.[241] Thus, the predictive value of legal data in the context of litigation will sometimes depend on information asymmetries. The degree of asymmetry will, in

---

240. *See supra* note 108 and accompanying text.

241. *See, e.g.*, Patrick J. Schiltz, *The Citation of Unpublished Opinions in the Federal Courts of Appeals*, 74 FORDHAM L. REV. 23, 78 (2005) ("In the past, some have also argued that, without no-citation rules, large institutional litigants (such as the Department of Justice) who can afford to collect and organize 'unpublished' opinions would have an unfair advantage."); Lauren K. Robel, *The Myth of the Disposable Opinion: Unpublished Opinions and Government Litigants in the United States Courts of Appeals*, 87 MICH. L. REV. 940, 955–59 (1989) (describing how repeat litigants gain an advantage over one-time litigants by circulating unpublished opinions among lawyers to help make decisions regarding case strategy, settlement, and whether to appeal).

some cases, determine whether the information remains advantageous for its users or becomes self-defeating over time.

Information asymmetries will likely remain prevalent given the high costs required to access, aggregate, process, and analyze large pools of litigation data.[242] The last two decades have already seen a widening gap in the financial status of firms and parties,[243] and this gap will translate into wide gaps in how much information each side is willing or able to purchase. To the extent that asymmetries prevail, the observations and predictions from big (legal) data will generally avoid becoming self-defeating or self-fulfilling and will instead merely work to the advantage of the side with better intelligence.

From a public policy standpoint, self-negation of information can produce a social good. For example, it can produce more consistency or uniformity between judges, which serves the goal of having similar litigants obtain equal results in every court. A judge may feel concern for his reputation if the data reveals a pattern of bias, delay, or haste, and the judge could accelerate the pace of self-correction of the prediction by modifying his own behavior. The information would self-correct eventually anyway, as parties will become aware of the bias and settle before trial or become aware of the delays or haste and seek other venues. Self-negation could also lead to earlier settlements where defendants determine, from an actuarial standpoint, that they stand to lose more by testing their luck at trial rather than settling the matter at the outset of litigation.

C. *Black Swans*

Surprise events can upset predictions, especially those based on the frequency of prior occurrences. Natural and man-made disasters, if they strike by surprise, can interrupt court processes, abruptly alter public opinion (thus impacting jurors, parties, and judges), and can disrupt markets (including currency values), which can change the stakes of a given case. Even personal events that come as a surprise—a judge or lawyer who has to step aside due to health problems—can nullify even a well-researched, well-analyzed prediction. Such "shocks" are ironically a familiar part of life, of course. The other side of the equation, however, is the tendency for predictions to foster systemic fragility. This is the core insight of bestselling author Nassim Nicholas Taleb in his trilogy on

---

242. *See supra* Part I.

243. Donald E. Vinson, *The Objective Edge: Third Party Assessment in Litigation Risk Analysis*, VINSON RESOLUTION MGMT., http://vinres.com/blog-2/litigation-risk-analysis/ (last visited July 1, 2015).

uncertainty—*Fooled by Randomness*,[244] *The Black Swan*,[245] and *Antifragile*[246]—that information abundance and the aggregate of predictive decisions can create an entirely new set of looming disasters, apart from what nature serves up.

Predictions and information present hazards in three ways: (1) decision makers tend to be overconfident when relying on forecasts, so they take more risks; (2) forecasters and decision makers tend to undervalue low-probability, high-stakes risks; and (3) over time, entire systems (institutional strategies and processes) grow up around the information so that sudden variations can cause exponentially greater catastrophes due to the increased interdependency within the entity.[247] Systemic efficiency—the elimination of redundancies, usually by centralized planning—makes systems more fragile, more like a house of cards. The insidious flipside of the "too big to fail" idea is that the same entities have become too big *not* to fail.[248] Taleb contrasts "fragile" institutions and individuals with those that are "robust"—independent enough to endure systemic shocks and unforeseen events.[249] Nature offers many examples of robustness—the ability of vegetation to regrow after defoliation, animal populations rebounding after an epidemic, and so forth. A third category—much rarer, and to Taleb, much more desirable—is what he dubs "antifragile," that is, individuals and entities that *benefit* from shocks of all types.[250] In the abstract, the legal profession is quite robust as evidenced by its antiquity. In the American context, the legal profession has even been rather antifragile, as Taleb would define the term, since sudden events causing social upheaval—catastrophic storms, wars, plane crashes, mass torts, crime sprees, market spikes or collapses, etc.—often create a gold rush for lawyers.

Yet the legal profession is changing in the age of big data. The availability of tools for harvesting insights from big (legal) data enables tighter coordination between parties, lawyers, and judges, because each of these players has access to far more information about the others, and can more accurately predict their responses to various courses of action. Reduced information costs, as a subset of transaction costs, foster more bargaining and yield earlier, more frequent agreements. Tighter

---

244. NASSIM NICHOLAS TALEB, FOOLED BY RANDOMNESS: THE HIDDEN ROLE OF CHANCE IN THE MARKETS AND IN LIFE (2001).

245. TALEB, THE BLACK SWAN, *supra* note 229.

246. NASSIM NICHOLAS TALEB, ANTIFRAGILE: THINGS THAT GAIN FROM DISORDER (2012) [hereinafter TALEB, ANTIFRAGILE].

247. *See* TALEB, THE BLACK SWAN, *supra* note 229, at 211.

248. *Cf.* Drury D. Stevenson & Nicholas J. Wagoner, *FCPA Sanctions: Too Big to Debar?*, 80 FORDHAM L. REV. 775, 778 (2011) (discussing the United States' refusal to debar a company because the company was "too big to debar").

249. *See* TALEB, ANTIFRAGILE, *supra* note 246, at 20–21.

250. *Id.* at 20–21, 85.

coordination means greater efficiency in that the reduction in information costs, as a subset of transaction costs, encourages bargaining in a greater number of cases and at an earlier stage in the pretrial litigation process. Thus, practitioners are beginning to step outside of the shadow of trial and are instead beginning to make important value judgments about their case in the shadow of data.

The downside is that tighter coordination raises the stakes for systemic errors or systemic effects from a breakdown in any part of the arrangement. This is not a reason to eschew big data, but rather an inherent feature of it that users should keep in mind. Taleb's main insight is not that unpredictable events happen, but that modern economic and political systems are more vulnerable to systemic shocks or disruptions because their tight coordination centers on precise, statistical predictions that allow no room for error, play, or disruption by a low-probability event.

Taleb's random events, "Black Swan" scenarios, and fragile systems are largely a response to the advent of big-data-based decision-making in the financial sectors (and also in the healthcare and educational systems, which have become increasingly complex but also increasingly interdependent). Simply put, data—especially statistical data—fosters interdependency in complex systems, and interdependency carries the inherent potential for system-wide disruptions or catastrophes. As the legal system begins to embrace the amazing potential of big data, these authors expect more interdependency and tighter coordination as statistical predictions furnish the basis for more decisions by lawyers, parties, and judges. Such increased interdependency, of course, increases the vulnerability to systemic disruptions or shocks. Again, this is not a reason to avoid big data, but it is something that sophisticated users of the data should recognize and for which they may want to compensate or hedge in some way.

D. *Ethical Considerations*

This Section discusses ethical issues that the rise of big data presents for legal practitioners and data-analytic firms. It starts with consideration of the prohibition against sharing fees with nonlawyers as implicated by the costs of parsing all of this data. It then addresses the potential big data has to create conflict-of-interest issues. Finally, it looks at the ways big data might change the lawyer's duty of care in the malpractice setting.

1. Unauthorized Practice of Law and Fee Sharing with Nonlawyer Owners

Does big data present new ethical quandaries or pitfalls for lawyers? Josh Blackman has raised several points that he believes present looming ethical problems for companies like Lex Machina that provide data

analytics to lawyers.[251] First, Blackman predicts that legal-data-analytic firms, which generally have lawyers and nonlawyers among their owners, will soon provide services that constitute the practice of law in violation of Rule 5.4 of the Model Rules of Professional Conduct, which bars fee sharing between lawyers and nonlawyers.[252] He writes that "[i]t is only a matter of time before several players in the field who are coy about the manner in which they are structured are going to have to deal with this [Rule]."[253]

Blackman's worries assume that data-analytic firms will: (1) engage in the practice of law as defined by state bars; (2) have nonlawyer owners when they do; and (3) base their operations outside Washington, D.C., which permits fee-sharing with nonlawyers.[254] At present, these authors are not aware of any data-analytic firms providing anything more than a sophisticated search engine for an online archive of legal information— no more the practice of law, say, than the familiar Westlaw or Lexis database services. Westlaw and Lexis both provide limited customer service or telephone assistance with searches, which to date have not been seriously challenged as the practice of law. Blackman seems to be suggesting that (in response to inevitable market demand) some data-analytic firms will go a step further and provide personal-consultant or data-concierge services through which a nonlawyer entrepreneur could obtain specific advice about where to file a lawsuit, whether to respond to a demand letter, and so on.[255] These authors are skeptical that the data-analytic firms will wade into such murky waters—their specialty, and the value-addition of their service, lies entirely in their database and search features, not with operating a clumsy legal-help call center where lawyers provide each caller with anonymous representation from within a cubicle. Nevertheless, concerns about variations on this idea are bound to arise, so it seems appropriate to work through the issues here for the sake of argument.

Blackman later raises the unauthorized practice of law (UPL) as a separate point: "Reading graphs to offer advice on how a case should settle, or where it should transfer to, is at its heart the practice of law. That an algorithm spit it out doesn't really matter."[256] As mentioned above,

---

251. Josh Blackman, *The Looming Ethical Issues for Legal Analytics*, JOSH BLACKMAN'S BLOG (Feb. 18, 2014), http://joshblackman.com/blog/2014/02/18/the-looming-ethical-issues-for-legal-analytics/.

252. *Id.*

253. *Id.*

254. *See* D.C. RULES OF PROF'L CONDUCT R. 5.4; ABA Comm. on Ethics and Prof'l Responsibility, Formal Op. 13-464 (2013).

255. *See id.*

256. *Id.*

defining the practice of law for disciplinary purposes is a necessary antecedent for his first point about sharing legal fees with nonlawyers, so it seems more appropriate to address these issues together. Applying the UPL issue to data analytics, Blackman asserts, "Non-lawyers, or even lawyers not working for a law firm, are unable to give this type of advice. Data-analytics firms should tread carefully about handing out this type of personalized advice outside the context of the attorney client relationship."[257]

The first step in responding to these concerns, therefore, is to address the issue of whether the activities performed by these data-analytic firms will fall within the definition of practicing law. Unfortunately, there is no clear, much less uniform, definition for this phrase. Few, if any, state bars have provided a clear definition of what constitutes "the practice of law," but certain actions have clearly fallen under this rubric—preparing and filing legal forms or pleadings and appearing as a client's representative before a tribunal are regularly featured.[258] The American Bar Association (ABA) concedes that there is no clear definition of the "practice of law," but it suggests functional proxies: "Although there is no single, specific definition of the practice of law, functionally the practice of law 'relates to the rendition of services for others that call for the professional judgment of a lawyer.'"[259] Merely providing descriptive information or analysis to lawyers or other customers, which is the current service provided by the data-analytic firms, is no more the practice of law than the algorithmic Google Trends or WestlawNext, albeit with richer data and interesting graphs and charts. If a data-analytic firm took the unusual step of drafting forms or filing documents on behalf of customers, outside the patent, trademark, Social Security, or immigration arenas, this would probably constitute representation and the provision of legal services.

The result, however, is less clear for the provision of consulting services. Jury consultants, who provide empirical-based advice about the likely biases or leanings of prospective or actual jurors, are pervasive in the legal system, though most are nonlawyer sociologists, psychologists, and so forth. Before Lex Machina or similar analytic firms begin offering services that verge on the practice of law, jury consultant firms are likely to use these resources to provide additional services to lawyers, and it seems unlikely that they would thereby trigger accusations of UPL. The ABA has issued no opinion letters about jury consultants, and there are no published opinions indicating that jury consultants have ever faced prosecution for UPL.

---

257. *Id.*

258. *See, e.g.*, Cleveland Bar Ass'n v. Boyd, 859 N.E.2d 930, 932 (Ohio 2006) ("With few exceptions . . . the unauthorized practice of law occurs when a layperson prepares legal pleadings and other papers for filing in court on another's behalf without the supervision of a licensed attorney.").

259. ABA Comm. on Ethics and Prof'l Responsibility, Informal Op. 1311 (1975).

In a published formal opinion in 2008, when the ABA approved of legal outsourcing by law firms to lawyers and nonlawyers, the ABA refused to define what might constitute UPL by an outsourced nonlawyer:

> Finally, the outsourcing lawyer must be mindful of the admonition of Rule 5.5(a) to avoid assisting others to "practice law in a jurisdiction in violation of the regulation of the legal profession in that jurisdiction . . . ." This Committee lacks the authority to express an opinion as to whether the provision of legal services by any particular lawyer, nonlawyer, or intermediary constitutes the unauthorized practice of law. Ordinarily, an individual who is not admitted to practice law in a particular jurisdiction may work for a lawyer who is so admitted, provided that the lawyer remains responsible for the work being performed and that the individual is not held out as being a duly admitted lawyer. We note only that if the activities of a lawyer, nonlawyer, or intermediary employed in an outsourcing capacity are held to be the unauthorized practice of law, and the outsourcing lawyer facilitated that violation of law by action or inaction, the outsourcing lawyer will have violated Rule 5.5(a).[260]

A survey of earlier ABA opinions is more illuminating. The historical trend has been to find improper fee sharing with lawyers—overlapped with UPL by the nonlawyers—almost exclusively in situations where the nonlawyer received contingent fees tied to the outcome of a case. For example, in a 1986 opinion about lawyers hiring outside researchers, the ABA forbade the payment of contingent fees to research corporations but did not forbid hiring them on an hourly basis:

> A corporation (the Corporation) is engaged in the business of providing legal research and analysis services to lawyers. In rendering these services the Corporation contracts only with lawyers. Its professional staff consists of law graduates, most of whom are licensed to practice in at least one American jurisdiction. The Corporation is not a professional law corporation and a substantial portion of its stock is owned by persons who are not lawyers. The Corporation has always billed on an hourly basis for its services to lawyers. It is now considering a contingent fee arrangement for situations where lawyers contemplate using the Corporation's services on a more extensive or continuing basis in the course of more complex matters, including complex litigation.[261]

---

260. ABA Comm. on Ethics and Prof'l Responsibility, Formal Op. 08-451 (2008).

261. ABA Comm. on Ethics and Prof'l Responsibility, Informal Op. 86-1519 (1986).

The ABA reached a similar conclusion in 1980 regarding lawyers who work for research firms that provide "sophisticated economic and economic-related analysis to enable lawyers to deal with problems for which such analysis is useful."[262] The owners of the research firm were lawyers and a nonlawyer economist.[263] The opinion described the services in terms that resemble those that Blackman contemplates the data-analytic firms will soon offer:

> This organization provides to lawyers advice on the most promising lines of economic argument, names of possible economic witnesses, consultation with witnesses, conducts economic studies to assist witnesses in testimony, and prepares comprehensive memoranda analyzing the facts as they bear on the economic issues in cases. . . . This consulting corporation, as described, does not preempt the attorney-client relationship. The final responsibility and supervisory powers remain with the attorney who hires this corporation. Your letter requesting a Committee opinion does not reveal in what manner the corporation is paid for its services to attorneys. In order to avoid ethical violations, the consulting corporation's compensation would have to be computed without regard to the legal fees collected by the attorneys requesting this service. Since employees of the corporation, or its agents, might be used as witnesses at trial[,] . . . it would be a violation of DR 7-109(C)(1)(3) to arrange a contingent fee agreement. Even if the consulting corporation had no witness role in a trial, a contingent fee agreement between attorney and his non-lawyer employees is forbidden . . . . [A]n accountant employee of a law firm must be "paid a regular salary computed without regard to fees collected for legal services rendered to particular clients." If the corporation is not practicing law, the monies paid to it by the employing attorney are not legal fees. Thus income can be ethically shared among lawyers and non-lawyers alike.[264]

The foregoing opinions illustrate how the definition of "practicing law" appears to blur with the analysis of whether fee sharing with nonlawyers occurs. The dispositive question thus appears to be whether the consulting firm bills its clients on a contingent fee basis—hourly or flat-fee billing appears to avoid any violations of the Model Rules.[265]

---

262. ABA Comm. on Ethics and Prof'l Responsibility, Informal Op. 1445 (1980).

263. *Id.*

264. *Id.*

265. On the other hand, ABA Committee on Ethics and Professional Responsibility, Informal Opinion 1264 (1973) dealt with a corporate-owned "Institute" that provided legal research for other attorneys:

Similarly, the ABA in 1987 took a dim view of "medical-legal consultants" primarily because the fee arrangement included contingent fees:

> The Consultant provides an initial "work product consultant report," usually through its "Medical Directors," consultation of its Medical Directors and, when a case so warrants, assistance to the lawyer with depositions of expert witnesses and opposing parties and at trial. It makes available expert witnesses from its "independent consulting staff." The Consultant offers services on a direct fee or contingent fee basis. The direct fee contract requires the client to pay agreed fees for certain services (including an "expert report charge" and a "testimony service charge" for each expert provided by the Consultant), reimbursable expenses, and all charges by the expert witnesses provided by Consultant (which are paid to the expert witnesses). The Consultant offers three contingent fee options: (1) a modified contingency fee of 20% of the total recovery under which the client pays a reduced report fee, reimbursable expenses and agreed reduced fees to expert witnesses; (2) a straight contingency fee of 30% of the total recovery under which the client pays only reimbursable expenses and no

---

[The Institute] advertises that it offers 'a full research organization for attorneys' promising a work product which will be presented in a 'creative, influential and convincing manner.' This service is stated by the advertisements to be available only to attorneys and to be performed by attorneys employed by The Institute. The Institute enters into contractual arrangements based upon an estimated fee with its customers at the outset of its service and bills them from time to time with the anticipation that it will make a profit. It, of course, compensates the attorneys working for The Institute in accordance with separate arrangements made with them. The Institute states it plans no contact with clients of individual attorneys and plans to limit its advertising to journals such as *The American Bar Association Journal* distributed primarily, but not exclusively, to members of the Bar.

The ABA Ethics Committee concluded that this was clearly unauthorized practice of law:

At the outset it must be pointed out that the corporation in question is not a professional association or corporation of lawyers, permissible under DR 2-102(B). Rather it is nothing more or less than a typical corporation for profit, utilizing lawyers for the performance of legal services for other lawyers. Ordinarily the definition of practice of law is a matter for which this Committee declines responsibility. However, we have little hesitancy in this instance in concluding that this is a classic case of an unauthorized practice of law and that lawyer participation is thus forbidden by DR 3-101(A).

Id.

fees for services of the Consultant or the expert witnesses provided by the Consultant; and (3) a fee to "maximize recovery" after a settlement offer which works like the straight contingency fee arrangement except that the percentage is negotiated and applies only to the amount of the recovery in excess of the settlement offer.[266]

At present, it seems that data-analytic firms will probably shield themselves from nonlawyer fee sharing or UPL accusations by marketing themselves to lawyers and law firms (including in-house counsel at corporations), refraining from offering case-specific strategic advice or predictions, and including disclaimers that their information does not constitute legal advice. Firms that want to branch out into case-specific consulting should probably package themselves similar to jury consultants like DecisionQuest, Malekpour & Ball, or High Impact, which tout themselves as full-service trial consultants.[267] If a new data-analytic startup were to base its headquarters in the District of Columbia (regardless of its server locations), law firms could share fees with them.[268] They would need to avoid charging contingent fees, though, as that has been the historical tripwire for the ABA to find outsourcing improper.

That nonlawyer ownership of law firms, within certain parameters, is permitted in the District of Columbia was recognized by another recent ABA Opinion:

> In contrast to the Model Rule, District of Columbia Rule 5.4(b) permits "an individual nonlawyer who performs professional services which assist the organization in providing legal services to clients" to hold an ownership interest in a law firm; and District of Columbia Rule 5.4(a) permits the sharing of legal fees with such persons . . . A fee-sharing issue may arise when a lawyer undertakes the representation of a client in a matter that involves the services of another lawyer or law firm governed by different rules. For example, a lawyer in a Model Rules jurisdiction may reasonably conclude that the client requires the assistance of a specific lawyer in a District of Columbia firm, in which a nonlawyer happens to hold an ownership interest, on a matter involving federal government contracts because that lawyer is uniquely qualified in such matters. With

---

266. ABA Comm. on Ethics and Prof'l. Responsibility, Formal Op. 87-354 (1987).

267. *See About Us: Overview*, DECISIONQUEST, http://www.decisionquest.com/Public/AboutUs/index.cfm (last visited July 1, 2015); *About M&B*, MALEKPOUR & BALL, http://malekpourball.com/ (last visited July 1, 2015); *About Us: Our Company*, HIGH IMPACT, http://highimpact.com/our-company (last visited July 1, 2015).

268. *See* D.C. RULES OF PROF'L CONDUCT R. 5.4; ABA Comm. on Ethics and Prof'l Responsibility, Formal Op. 13-464 (2013).

informed client consent, the two lawyers may work together on the matter. If the requirements of Model Rule 1.5(e) are met, a typical fee arrangement in such matters is for the client to receive a single billing for the work of both lawyers. In this situation, there may be a question whether the lawyer from the Model Rules jurisdiction, by participating in this common inter-firm fee arrangement, shares a legal fee in violation of Model Rule 5.4(a) because the District of Columbia firm's portion of the fee will presumably become part of that firm's overall revenues, revenues from which distributions may ultimately be made to the nonlawyer who holds an ownership interest.[269]

If a data-analytic firm with nonlawyer owners was to base its operations in Washington, D.C., law firms in other jurisdictions could share fees with the lawyers employed by the data firm.[270] Given that their services are entirely online (with some additional phone service), there is no reason that they could not incorporate and register their business in D.C. as opposed to Silicon Valley.

It is also worth mentioning that there is wide variation in how states enforce the rules against UPL. For example, a number of states have an Unauthorized Practice of Law Commission that functions somewhat separately from the state bar disciplinary authority.[271] The Commission may even have a different definition of UPL than the state entity that promulgates ethics opinions or the entity that reprimands or disbars lawyers.

2. Conflicts of Interest

Next, Blackman identified conflicts of interest as a likely problem for big data: "[T]hey will need to put safeguards into place to ensure that they are not offering advice to adverse parties, or more likely, clients who may

---

269. ABA Comm. on Ethics and Prof'l Responsibility, Formal Op. 13-464 (2013).

270. *See id.*

> Lawyers subject to the Model Rules may work with other lawyers or law firms practicing in jurisdictions [such as the District of Columbia] with rules that permit sharing legal fees with nonlawyers. Where there is a single billing to a client in such situations, a lawyer subject to the Model Rules may divide a legal fee with a lawyer or law firm in the other jurisdiction, even if the other lawyer or law firm might eventually distribute some portion of the fee to a nonlawyer, provided that there is no interference with the lawyer's independent professional judgment.

Id.

271. *See, e.g., About Us*, TEXAS UPL COMMITTEE, http://www.txuplc.org/Home/about (last visited July 1, 2015).

have interests adverse to another client."[272] To the extent that these data firms have outside funders or venture capitalists, this would further complicate the range of potential conflicts of interest.[273] Moreover, Blackman expects that the data firms will want to harvest data from their clients, as well as provide information, in order to enrich their database and enhance their algorithms.[274] Blackman anticipates that users of these services would find it objectionable that their search trails would be visible—even partly or indirectly—to litigation opponents.[275]

Presumably, user history will be invisible to other users, or invisibility or anonymous searching will at least be an option for users, as is the case with online sites like YouTube and LinkedIn. Even so, user-informed algorithms produce results from which sophisticated users could, in theory, infer information about other users, especially when one's litigation opponent is the only other user likely to have searched for information about one's firm, one's client, the judge assigned to the case, and so on. Most cases, even for seasoned litigators, have unique combinations of parties, counsel, experts, and judges. Just as Google's autocomplete feature in its search box reveals information about the most common searches around the world beginning with the letters typed so far, if a data-analytic site were to have a similar autocomplete feature, it could reveal search queries by an opposing party.

In response to these concerns, this Article first offers the observation that this concern raised by Blackman runs counter to his concerns about sharing fees with nonlawyers and UPL because the conflicts of interest rules will not apply unless the data firms are, or include, lawyers engaged in the practice of law. Conversely, if a big data firm turns out to be lawyers providing legal, but mostly statistics-based, services for clients, rules about conflicts would apply but concerns about UPL would not. Sharing legal fees with nonlawyers would also be less likely problematic.

Moving on to the rules about conflicts of interest, the ABA Model Rules of Professional Conduct cover conflicts of interest starting with the basics at Rule 1.7, then address special types of conflicts of interest in Rule 1.8, including the use of "information relating to representation of a client to the disadvantage of the client."[276] Both Rules 1.7 and 1.8, however, have consent-waiver provisions with low thresholds.[277] Data-analytic firms providing services to lawyers will presumably obtain informed consent from users as part of their boilerplate terms and conditions of use. The services, as they function now, present no more

---

272. Blackman, *supra* note 251.
273. *Id.*
274. *Id.*
275. *See id.*
276. MODEL RULES OF PROF'L. CONDUCT R. 1.8(b).
277. *See* MODEL RULES OF PROF'L. CONDUCT R. 1.7, 1.8.

conflict of interest than would Westlaw or Lexis providing search results to litigation opponents who use their services at the same time. Even if these firms eventually offer concierge or individualized consulting services, the potential conflicts implicated are not any greater than a jury consultant who represents different firms and who at times has been on different sides of litigation or transactions regarding certain parties. Rule 1.8(b)'s prohibition applies only to the use of information that disadvantages a client,[278] so non-injurious use of data harvested by the entity from its clients would not necessarily trigger a violation of Rule 1.8 unless the client could demonstrate harm, which seems unimaginable.

3. Legal Malpractice

Blackman's third ethical concern pertains to legal malpractice insurance.[279] He offers a hypothetical of a big data firm that advises a user to file suit in a particular forum or venue based on the favorable patterns there, but fails to account for a recent change in the law in that jurisdiction, which ensures an unfavorable outcome from now on.[280] Blackman sees this as a potential malpractice suit, and he expresses concern that data-analytic firms will not carry malpractice insurance.[281] He writes that "[i]n most jurisdictions, active lawyers are required to maintain some level of malpractice insurance to protect against this."[282]

Actually, only one state—Oregon—requires lawyers to carry legal malpractice insurance.[283] A few states require lawyers to provide notice

---

278. MODEL RULES OF PROF'L. CONDUCT R. 1.8(b).

279. *See* Blackman, *supra* note 251.

280. *Id.*

281. *See id.*

282. *Id.*

283. Jack A. Guttenberg, *Practicing Law in the Twenty-First Century in a Twentieth (Nineteenth) Century Straightjacket: Something Has to Give*, 2012 MICH. ST. L. REV. 415, 472 n.339 ("Oregon is the only American jurisdiction to require malpractice insurance of all lawyers who are admitted to practice in the state. As of Aug. 9, 2011, seven jurisdictions require disclosure directly to the client of the lawyer's failure to maintain malpractice insurance. Another seventeen jurisdictions require lawyers to report their lack of malpractice insurance on their annual registration statements." (citation omitted)); Susan Saab Fortney, *Seeking Shelter in the Minefield of Unintended Consequences—The Traps of Limited Liability Law Firms*, 54 WASH. & LEE L. REV. 717, 729–30 (1997) ("Some LLP and LLC statutes provide that the liability shield only applies if the LLP or LLC carries some requisite amount of insurance coverage. Other statutes do not condition liability protection on maintaining some level of insurance or providing evidence of financial responsibility." (footnote omitted)); Leslie C. Levin, *Bad Apples, Bad Lawyers or Bad Decisionmaking: Lessons from Psychology and from Lawyers in the Dock*, 22 GEO. J. LEGAL ETHICS 1549, 1588 (2009) ("[L]egal malpractice insurance is not required in the United States in order to practice law."); Martin C. McWilliams Jr., *Who Bears the Costs of Lawyers' Mistakes?— Against Limited Liability*, 36 ARIZ. ST. L.J. 885, 948 (2004) ("As of this writing only one state requires lawyers to carry insurance, and in other states movements to require legal malpractice

to clients that the lawyer is uninsured at the beginning of representation—a mere formality, fine print in the retainer agreement.[284] And a few states require all LLCs and LLPs—not only lawyers, but certainly including them—to carry a minimum amount of liability insurance, usually $100,000.[285] Thus, firms using the corporate form of LLC or LLP in those states are required to have liability insurance, but not necessarily legal malpractice insurance in particular. In general, therefore, lawyers in forty-nine states do not have to carry professional malpractice insurance, thus making the data-analytic firms no different than most lawyers in regard to required malpractice insurance coverage.

According to a 2013 survey of the seven largest legal malpractice insurers in the country, conflicts of interest generated the lion's share of malpractice claims, with "procedural errors" and "failure to know/properly apply law" tied as distant seconds (only one-fourth as common as conflicts as the cause).[286] The category of "litigation errors" (it is unclear how this was distinguished from "procedural errors") was the least common cause of malpractice claims that the survey actually reported.[287]

Returning to Blackman's hypothetical in which a nonlawyer consultant gives misinformed advice about where to file a suit,[288] liability for malpractice, again, depends heavily on the firm meeting some definition of the practice of law. There is no clear definition—and legal malpractice cases against nonlawyers are extraordinarily rare. This scenario is analogous to legal malpractice suits against jury consultants—again, rare or nonexistent. Moreover, a legal malpractice action typically requires the plaintiff to prove damages and causation—that the case could have prevailed but for the defendant's negligence. It is not clear under Blackman's example that the party could not refile the case in another jurisdiction, thus resolving the harm and undermining a malpractice claim.

---

insurance have failed." (footnote omitted)); Carole Silver, *Regulating International Lawyers: The Legal Consultant Rules*, 27 HOUS. J. INT'L L. 527, 535 (2005) (noting that seventeen states require foreign legal consultants to have proof of insurance to be licensed, but only Oregon requires lawyers to carry malpractice insurance); Christel Walther, *LLC and Lawyers: A Good Combination?*, 50 LOY. L. REV. 359, 390–91 (2004).

284. Walther, *supra* note 283, at 392–93.

285. *Id*. at 391 & n.205, 392–93.

286. AMES & GOUGH, LAWYERS' PROFESSIONAL LIABILITY CLAIMS TRENDS: 2013, at 6 (2013), *available at* https://www.estateworks.com/home/marketingsite/data/AmesGough2013.pdf (insurer survey).

287. *Id.*

288. Blackman, *supra* note 251.

E. *Privacy*

Recent academic commentary about big data has focused disproportionately on privacy concerns,[289] and these authors hope to turn the discussion in another direction. This Article does not offer a full-blown rebuttal to the alarmist, Orwellian predictions of others, as that would turn this Article into yet another piece focused on an overdone subject. It seems appropriate to provide a brief justification for the authors' call to move beyond the excessive hand-wringing about big data and personal privacy.[290]

Academic fretting about privacy began midcentury as a legitimate reaction against McCarthyism,[291] but became overblown in the 1980's and 1990's as an artifact of the narcissism and hyper-individualism of the culture in that era. The endless dire predictions that this country was on the verge of totalitarianism, or an American version of the Night of the Long Knives,[292] now seem like a vestigial pushback against the Red Scare of the 1950's. The much-feared rounding up of political dissidents, freethinkers, or sexual experimenters never materialized again. The public discussion of privacy rights has evolved, of course. Now writers seem concerned not only with buffering the power of a modernized surveillance state, but also with the perceived information asymmetries between consumers and marketers—the "big corporations" that will supposedly injure our society with targeted advertising. There is no need to add to the growing supply of academic writing about the evils of companies showing citizens the products and services most relevant for their age group, gender, and geography.

A similar strain of alarmist literature centers on a phobia that societal taboos will be easier to enforce if people cannot hide their vices. Yet this is not as much of a problem with privacy as it is with the fact that this country's taboos are changing, so there are divergent sets of values or taboos that groups seek to impose on each other. There are emerging taboos against racism, misogyny, and other forms of bigotry; greater transparency would help new norms of tolerance to emerge.

The second reason for moving beyond the tired privacy discussion flows from this last point about the new values in this society. This Article suggests that the country cannot pull together its balkanized factions and newly-arrived subcultures without greater transparency—anonymity and community are inherently in tension.

---

289. *See supra* note 189 and accompanying text.

290. For a more thorough rebuttal to scholars such as Julie Cohen, see Jane Bambauer, *Is Data Speech?*, 66 STAN. L. REV. 57, 96 (2014).

291. *See, e.g.*, George Anastaplo, *"McCarthyism," the Cold War, and Their Aftermath*, 43 S.D. L. REV. 103, 103–04 (1998).

292. *See* Telford Taylor, *The Nuremberg Trials*, 55 COLUM L. REV. 488, 513 (1955) (noting the history of the Night of the Long Knives).

CONCLUSION

A seismic shift is underway in the legal profession, whereby lawyers are increasingly supplementing their intuition and practice experience with insights gleaned from big (legal) data to inform their judgment. The increasing accessibility of legal data and continued refinement of the analytics for mining it will alter the skillset and approach of lawyers who embrace these tools. The analytics platforms that are facilitating this shift will not, however, supplant practicing attorneys; rather, these platforms will complement practitioners' skills, knowledge, and experience. While algorithms may be able to process and summarize trends and patterns in historical litigation data far more accurately and efficiently than their human counterparts, it is ultimately the lawyer's job to draw on her experience and intuition to determine which trends and patterns are relevant to a given situation and how to act on such insight.

Even so, the point here has not been to promote technophilia. Tempering exuberance for new predictive tools should be a more reflective, nuanced approach to predictions generally—a recognition that predictions can backfire, become self-fulfilling, or can even create new systemic risks or institutional fragility. Knowing the limitations of a method or technology ultimately makes it more useful. Risks and recklessness are not necessarily the same.

This Article's discussion has therefore been primarily descriptive but also modestly normative. The descriptive portions explained the new platforms that are available for legal data analysis, and the normative portions explored the pros and cons of using these tools in practice. Most advances in technology merely made lawyers more efficient at doing the same tasks they were already doing. Big data changes the nature of making legal predictions—and thereby promises to change the role of lawyers overall.