

# Citation Stickiness, Computer-Assisted Legal Research, and the Universe of Thinkable Thoughts

Aaron S. Kirschenfeld\*

Alexa Z. Chew\*\*

## Introduction

This article seeks to answer two main questions. The first is whether courts cited the same cases as the parties more often during the print era than during the digital era. The second is what, if anything, the answer to the first question can contribute to the debate about how print-era forms of organizing and describing case law influenced researchers' behavior. To that end, we sampled cases from 1957, 1987, and 2017, and used "citation stickiness" to study the differences in how parties and judges cited authorities during each of those years. In short, we found that there is less agreement about what case law authorities are relevant to an appeal between parties and judges in 1957 than in 1987 and 2017. This casts doubt on the existence of a cozy "universe of thinkable thoughts," or the longstanding theory that classification schemes like West's American Digest System led to greater coherence and stability in the development of common law in the United States.

In section I of this article, we review the literature on how switching from print research to digital research influences lawyers' research habits and conceptions of the law. We then look at prior empirical studies assessing the kind of law found by researchers within different research environments or by using different research processes.

---

\* Digital Initiatives Law Librarian and Clinical Associate Professor of Law, University of North Carolina at Chapel Hill.

\*\* Clinical Professor of Law, University of North Carolina at Chapel Hill, who had top notch help from her UNC Law research assistants, Taylor Carrere and Marshall Newman. This study would not have been possible without their careful work.

In section II of this article, we introduce the citation stickiness metric and describe our methodology.

In section III we present our results, which show that there is a significant difference between 1957, 1987, and 2017 in how often courts cite cases originally cited in at least one party's brief. We also explore some other possible conclusions gleaned from our data. Finally, we speculate on the reasons why we found what we found and identify questions for further study.

## I. "Thinkable thoughts" and legal research

This section considers the issues raised by the vibrant and long-standing debate over the influence of print-era case law classification systems on legal research and the development of common law in the United States.

### A. The influences of print-era case law classification systems

Did tools developed during the print era to publish, describe, and classify case law also influence the ways lawyers thought about the law and, consequently, the way that law developed? Many law librarians and legal scholars have taken up this question in the past forty years.<sup>1</sup> Some have contended that the American Digest System had a good deal of influence.<sup>2</sup> Some, less so.<sup>3</sup>

The arguments advanced are complex, but for our purposes can be reasonably simplified as follows: print-era classification systems and patterns of publication created coherence and stability in the landscape of legal information. Early digital sources mirrored the structure of these systems and patterns of publication, but new tools and sources made available during the digital era would challenge the ways that researchers come to know law.

The work of Bob Berring deserves special attention.<sup>4</sup> It posits that tools like the American Digest System and the headnotes that constituted it normalized "legal language and legal meanings . . . [forming] the ground

<sup>1</sup> Stefan H. Krieger & Katrina Fischer Kuh, *Accessing Law: An Empirical Study Exploring the Influence of Legal Research Medium*, 16 VAND. J. ENT. & TECH. L. 757, 759 n.6 (2014) (collecting articles and studies about "the influence of digitization on the law generally and on legal research specifically").

<sup>2</sup> See Richard A. Danner, *Influences of the Digest Classification System: What Can We Know?*, 33 LEGAL REFERENCE SERVS. Q. 117, 128 n.49 (2014) (collecting works about "the extent of the digest's influences in categorical terms").

<sup>3</sup> See, e.g., Peter C. Schanck, *Taking Up Barkan's Challenge: Looking at the Judicial Process and Legal Research*, 82 LAW LIBR. J. 1 (1990).

<sup>4</sup> For an excellent summary of Berring's work on this topic, see Richard A. Danner, *Legal Information and the Development of American Law: Writings on the Form and Structure of the Published Law*, 99 LAW LIBR. J. 193 (2007).

of integration and coherence in substantive law.” These tools, in turn, influenced “the way legal researchers conceptualized the law.”<sup>5</sup> Indeed, the classification systems and publishing patterns created “a cozy universe” of legal meaning such that “all of those trained within it have created a conceptual universe of thinkable thoughts that has enormous power.”<sup>6</sup>

Plenty of other scholars have addressed questions about “the extent to which the Key Number System influences the law itself.”<sup>7</sup> Barbara Bintliff described the West digests as “allowing researchers to understand the relationship, context, and hierarchy of identified rules . . . . [Lawyers] have to think in terms that match its organization.”<sup>8</sup> Before the advent of computer-assisted legal research (CALR), digests and “a predictable, stable judicial system . . . became almost inextricably intertwined.”<sup>9</sup> Bintliff noted the difficulties of constructing computerized systems that would allow researchers to discover legal rules as readily as was possible in the print era, but left the door open to technological advances someday catching up.<sup>10</sup> More on that in a moment.

Carol Bast and Ransford Pyle added to that line of thinking with words of further warning.<sup>11</sup> They described the move to CALR as a paradigm shift away from coherence and stability in the law, legal thinking and, by extension, legal research.<sup>12</sup> The paper concluded that digital resources and processes will bring about “a more primitive legal regime,” lessening lawyers’ consensus understanding of hierarchic legal concepts.<sup>13</sup>

F. Allan Hanson then added an anthropological perspective to this argument in his analysis of information management systems and the law.<sup>14</sup> In seeking to explain what differentiated print resources and

<sup>5</sup> Robert C. Berring, *Legal Research and Legal Concepts: Where Form Molds Substance*, 75 CALIF. L. REV. 15, 22 (1987); see also Robert C. Berring, *Ring Dang Doo*, 1 GREEN BAG 2D 3, 3 (1997) (“Without realizing it, we all depended on West for giving us ways to think coherently about the hundreds of thousands of cases that were stuffed into the reporters.”).

<sup>6</sup> Robert C. Berring, *Legal Research and the World of Thinkable Thoughts*, 2 J. APP. PRAC. & PROCESS 305, 311 (2000) [hereinafter Berring, *World of Thinkable Thoughts*]; see also Robert C. Berring, *Legal Information and the Search for Cognitive Authority*, 88 CALIF. L. REV. 1673, 1693 (2000) [hereinafter Berring, *Search for Cognitive Authority*] (“Generations of lawyers learned to conceptualize legal problems using the categories of the Topics and Key Numbers of the American Digest System.”).

<sup>7</sup> Daniel Dabney, *The Universe of Thinkable Thoughts: Literary Warrant and West’s Key Number System*, 99 LAW LIBR. J. 229, 230 (2007).

<sup>8</sup> Barbara Bintliff, *From Creativity to Computerese: Thinking Like a Lawyer in the Computer Age*, 88 LAW LIBR. J. 338, 343 (1996).

<sup>9</sup> *Id.* at 344.

<sup>10</sup> *Id.* at 351.

<sup>11</sup> Carol M. Bast & Ransford C. Pyle, *Legal Research in the Computer Age: A Paradigm Shift?*, 93 LAW LIBR. J. 285 (2001).

<sup>12</sup> *Id.* at 286.

<sup>13</sup> *Id.* at 302.

<sup>14</sup> F. Allan Hanson, *From Key Numbers to Keywords: How Automation Has Transformed the Law*, 94 LAW LIBR. J. 563 (2002).

research processes, he also focused on the “hierarchical, taxonomic classification” of the digests, arguing that their categories “have been reified into principles thought to preside over ‘the law,’ understood as a self-contained, independently existing system.”<sup>15</sup> Indeed, Hanson saw automated research as a threat to the doctrine of precedent, a cornerstone of the common law.<sup>16</sup> As for research, computerized systems were apt to turn up a wider variety of cases that could be considered precedential, unlike in the print era, when “opposing attorneys would tend to develop their arguments on the basis of the same cases, nearly all of which were familiar to judges and experts in that field of law.”<sup>17</sup>

Jean Stefancic and critical race theory co-founder Richard Delgado also weighed in, arguing that “professionally prepared research and indexing systems . . . function like DNA; they enable the current system to replicate itself endlessly, easily, and painlessly.”<sup>18</sup> And in doing so, these print-era systems facilitate the quick research of traditional legal arguments but hamper the research needed for innovative jurisprudence.<sup>19</sup> Writing “at the dawn of the computer revolution”<sup>20</sup> in 1989, Delgado and Stefancic opined that “[c]omputerized word-search strategies promise some hope of breaking the constraints imposed by older systems” by, for example, allowing a researcher to “combin[e] two [index] categories in the same search.”<sup>21</sup> This hope had largely dissipated when Delgado and Stefancic revisited their triple helix dilemma in 2007: “our predicament is little better than it was in the days of searching in the dusty volumes of the West decennial digests and, in some respects, more acute.”<sup>22</sup> They argued that CALR “may in fact impede the search for new legal ideas” in part because legal training still taught lawyers to think in terms of print-era index categories.<sup>23</sup>

<sup>15</sup> *Id.* at 570.

<sup>16</sup> *Id.* at 579.

<sup>17</sup> *Id.* at 580 (citing Bintliff, *supra* note 8, at 343–44). Bintliff’s claim, upon which Hanson’s relies, is a descriptive one: “Lawyers in Florida and South Dakota, Ohio and Nevada, consulted the same books, used the same organizing framework, found the same cases. The arguments crafted from these cases encouraged the best legal thinking, and gave judges the opportunity to explore the many sides of an issue and make a decision that was understandable.” However, Bintliff’s descriptive claim is not obviously supported in *Thinking Like a Lawyer*. It appears to be a “common sense” claim rather than one supported by historical research.

<sup>18</sup> Richard Delgado & Jean Stefancic, *Why Do We Tell the Same Stories?: Law Reform, Critical Librarianship, and the Triple Helix Dilemma*, 42 STAN. L. REV. 207, 208 (1989).

<sup>19</sup> *Id.*

<sup>20</sup> Richard Delgado & Jean Stefancic, *Why Do We Ask the Same Questions—The Triple Helix Dilemma Revisited*, 99 LAW LIBR. J. 307, 309 (2007).

<sup>21</sup> Delgado & Stefancic, *supra* note 18, at 209, 219.

<sup>22</sup> Delgado & Stefancic, *supra* note 20, at 310.

<sup>23</sup> *Id.* at 310.

These arguments about revolutionary changes can be understood in context, as legal publishers consolidated, added materials, and developed newer and more powerful computerized resources. In 1986 and 1987, when Berring began writing on the topic, full-text searching on Westlaw had only been available for a handful of years. And throughout the 1990s, the habits of law students were changing. On both counts, it was certainly worth speculating—and cautioning—about the new era to come.<sup>24</sup> But after a decade of relative stasis in how legal resources are created and how legal research is conducted, we figured it was time for a reappraisal.

We have chosen to pick up this line of thought with a question posed by Dick Danner: “How does one show what influences research tools might have on lawyers’ thinking about the law . . . during the late twentieth century when print digests began to be bypassed in favor of electronic tools?”<sup>25</sup> The first step in that process is to look at studies that have attempted to answer it.

## B. Studies of research and resources

In seeking to quantify the influence of print-era classification tools on the habits of legal researchers, law librarians and other legal scholars have conducted surveys and crunched numbers. There have been many empirical studies on the topic.<sup>26</sup> Below, we look at the ones relevant to our question.

### 1. User studies

Several studies have looked at the thought processes and habits of legal researchers to distinguish between how researchers use print or print-era sources and how they use electronic sources and methods.

Lee Peoples set about to test whether researchers use digests or other subject-organized systems to locate relevant legal rules but use electronic sources, such as full-text searches, to locate relevant facts.<sup>27</sup> To that end, Peoples designed a study to learn whether electronic resources were superior to print digests for locating cases with similar fact patterns.<sup>28</sup> The subjects were law students, and the study was conducted in 2004.<sup>29</sup>

---

<sup>24</sup> Berring noted that the changes he had anticipated in the late 1980s were not as extensive as he had suspected. Berring, *Search for Cognitive Authority*, *supra* note 6, 1707–08; see also Hanson, *supra* note 14, at 579.

<sup>25</sup> Danner, *supra* note 2, at 129.

<sup>26</sup> *Id.* at 134 n.74.

<sup>27</sup> Lee F. Peoples, *The Death of the Digest and the Pitfalls of Electronic Research: What Is the Modern Legal Researcher to Do?*, 97 *LAW LIBR. J.* 661 (2005).

<sup>28</sup> *Id.* at 668.

<sup>29</sup> *Id.*

Peoples's results cast doubt on the hypothesis that students would be more successful locating relevant legal rules by using digests and more successful locating relevant fact patterns by using full-text searching.<sup>30</sup> The study found that electronic resources were not superior to print digests for finding cases with similar fact patterns.<sup>31</sup> Electronic sources, likewise, were not superior to print sources for locating relevant legal rules.<sup>32</sup>

This study is important to our inquiry because it challenges the notion, developed in the literature, that print-era tools would be better for locating relevant legal rules. It suggests that the structure of these print-era tools may not be as influential on the thoughts and habits of legal researchers as theorized. But there are a couple of problems. First, 2004 is far enough in time from the introduction of electronic sources that differences between print-era structures and digital structures may be hard to parse. Second, the subjects of the study were law students, who might be presumed to have less experience with solving legal problems and using legal sources than practicing attorneys.

A few years later, the behavior of practicing attorneys was studied by Joseph Custer, who cast a bit more light on how researchers with more domain-specific problem-solving experience would use legal resources.<sup>33</sup> Custer's survey sought to test whether (1) attorneys use more than one system to locate relevant law, (2) some attorneys never use digests, (3) attorneys tend to research facts more than legal rules or doctrines, and (4) attorneys pay little attention to digest categories.<sup>34</sup> The subjects of the study were attorneys in Kansas.<sup>35</sup>

Significantly for us, the survey found that more than half of the attorneys did not use digests at all.<sup>36</sup> It also found that attorneys pay little attention to digest categories.<sup>37</sup> These findings challenged assertions that print-era digest categories led attorney researchers to think about the law in terms of those abstract classifications.<sup>38</sup> Instead, the findings suggest a weak connection between practitioners and subject-based classifications of case law. However, the survey was conducted in the late aughts, meaning it is even further in time from the introduction of digital

<sup>30</sup> *Id.* at 670.

<sup>31</sup> *Id.*

<sup>32</sup> *Id.*

<sup>33</sup> Joseph A. Custer, *The Universe of Thinkable Thoughts Versus the Facts of Empirical Research*, 102 LAW LIBR. J. 251 (2010).

<sup>34</sup> *Id.* at 258. The contentions were derived from those first posed by Schanck, *supra* note 3.

<sup>35</sup> *Id.*

<sup>36</sup> *Id.* at 260.

<sup>37</sup> *Id.* at 262–63.

<sup>38</sup> *Id.* at 264. Custer's criticism is mostly directed at Dabney, *supra* note 7.

sources than Peoples's study. Again, we were stymied in our search for a study comparing how practitioners would behave with print-era sources as compared with digital sources.

Susan Nevelow Mart tested the differences between subject-organized case law systems created with human intervention and those created with computer algorithms.<sup>39</sup> The study pitted subject-organized system against subject-organized system, and it found that researchers were more successful using systems where case indexing was done by humans.<sup>40</sup> The research subjects were law students, and the study was conducted in the early 2010s.<sup>41</sup>

Finding that a higher percentage of relevant cases are located using a human-curated case-indexing system<sup>42</sup> suggests that the print-era digest systems remained powerful tools for researchers looking to find legal rules well into the electronic era. The question remains, however, whether print-era tools would perform the same way when they were the only game in town.

Stefan Krieger and Katrina Fischer Kuh sought to study the differences between the processes used in print and electronic research, as well as the results of each.<sup>43</sup> Law students in the early 2010s were the subjects, and these students researched a problem and described their research processes.<sup>44</sup> Half used print sources and half used electronic sources.<sup>45</sup>

The study's findings showed that students conceived of and structured their research differently depending on which research medium they were using.<sup>46</sup> The findings are at odds with those of Custer, suggesting that "electronic researchers can, in fact, be expected to emphasize fact terms as compared to legal concepts in their research and to rely more on primary sources and less on secondary sources than print researchers."<sup>47</sup> This tension might be the result of the different populations studied by each, or perhaps of the small sample size used by Krieger and Kuh. It also might be the result of different legal research training. The study subjects were

---

<sup>39</sup> Susan Nevelow Mart, *The Case for Curation: The Relevance of Digest and Citator Results in Westlaw and Lexis*, 32 LEGAL REFERENCE SERVS. Q. 13 (2013).

<sup>40</sup> *Id.* at 14–15.

<sup>41</sup> *Id.* at 26.

<sup>42</sup> *Id.* at 38.

<sup>43</sup> Krieger & Kuh, *supra* note 1, at 762.

<sup>44</sup> *Id.* at 766–67.

<sup>45</sup> *Id.* at 762.

<sup>46</sup> *Id.*

<sup>47</sup> *Id.* at 789.

selected in part based on the students' print research experience beyond the required first-year course, which typically included one print research assignment,<sup>48</sup> but the study does not describe how first-year students were instructed to use print sources and electronic sources. In any event, whether a tendency to focus on facts in research using electronic sources suggests much of anything about the influence of print-era systems also remains unanswered.

These studies, conducted on users of legal research systems, approach the problems raised by Berring from different angles and ultimately do not reach a consensus. None test attorney research habits from the print era. To get a better sense of that, we turn now to citation studies that more directly address the question.

## 2. Citation studies

Two recent studies sought to explain historical differences between pre-CALR legal research and post-CALR legal research. Both, like ours, are citation studies of court decisions. And both, therefore, consider the work of practitioners—namely, judges—and draw data from the past. But both studies also limited their scope to judicial writing, looking at citation practices in judicial opinions but not attorneys' briefs.

Paul Hellyer studied a sample of California Supreme Court opinions to test whether research is more efficient using CALR tools and whether those tools reshape the law.<sup>49</sup> Looking at a sample of 180 cases from 1944 to 2003, Hellyer sought to identify changes in quantity, recency, and type of legal authority cited by courts.<sup>50</sup> Hellyer hypothesized that, if CALR had influenced research practices, contemporary courts would be (1) citing more cases in their opinions, (2) citing more cases from outside their jurisdiction, (3) citing more recent cases, (4) citing authorities only available electronically, and (5) citing more secondary sources as authoritative.<sup>51</sup> Hellyer did find "some significant changes in the court's citations to legal authority," but concluded that there was "no clear indication" that the introduction of CALR had caused the changes.<sup>52</sup>

Hellyer's study differs from ours in several important respects. First, it studied only judicial behavior, and judges form only a small subset of all practitioners. Second, it studied only citations in majority opinions,<sup>53</sup>

---

<sup>48</sup> *Id.* at 764 n.29.

<sup>49</sup> Paul Hellyer, *Assessing the Influence of Computer-Assisted Legal Research: A Study of California Supreme Court Opinions*, 97 *LAW LIBR. J.* 285 (2005).

<sup>50</sup> *Id.* at 285.

<sup>51</sup> *Id.* at 290.

<sup>52</sup> *Id.* at 293.

<sup>53</sup> *Id.*

thus excluding citations in concurrences and dissents that might reflect additional judicial research. Third, it excluded citations that appeared in quotations from other cases and citations to prior opinions in the same case, which would lead to a lower number of cited cases than our study. And finally, by analyzing only three cases per year, the results are likely difficult to replicate.

Next, Casey Fronk conducted an empirical analysis of 1,200 federal appellate cases from 1957 to 2007.<sup>54</sup> The study was designed, among other things, to examine “quantitative and stylistic” changes in judicial citation practices resulting from changing research sources.<sup>55</sup> Like our study and unlike Hellyer’s, Fronk’s methodology relied on Westlaw’s “Table of Authorities” feature,<sup>56</sup> and therefore included all unique case citations in majority, concurring, and dissenting opinions.

Fronk found the greatest effect of computerized legal research on judicial citation practices between 1977 and 1987.<sup>57</sup> This conclusion was reached by showing the growth of expository citation over string citation as access to CALR increased.<sup>58</sup> While useful in terms of showing both the quantitative and qualitative changes in judicial citation, the study does not examine changes in how advocates, more broadly, have conducted legal research over time.

Hellyer’s study concluded by saying that “CALR’s effects on courts cannot be measured by an analysis of citations in court opinions. If this is true, what is the appropriate measurement?”<sup>59</sup> We think we have an answer.

## II. Measuring citation stickiness

Next, we introduce the citation stickiness metric and describe the results of an initial study of the topic, and why it is useful for exploring the concepts of legal information discovery tools and their influence on interpreting law.

---

<sup>54</sup> Casey R. Fronk, *The Cost of Judicial Citation: An Empirical Investigation of Citation Practices in the Federal Appellate Courts*, 2010 U. ILL. J.L. TECH. & POL’Y 51, 53 (2010).

<sup>55</sup> *Id.* at 53, 67.

<sup>56</sup> *Id.* at 67–68.

<sup>57</sup> *Id.* at 78. Fronk describes the result of a 1976 study of “actual federal court research methods” that showed that federal appellate law clerks used CALR systems from 0.26 to 7.33 hours per month, and that monthly usage by district court law clerks was less than half that. *Id.* at 61 (summarizing ALAN M. SAGER, AN EVALUATION OF COMPUTER ASSISTED LEGAL RESEARCH SYSTEMS FOR FEDERAL COURT APPLICATIONS 77 tbl.25 (1977)).

<sup>58</sup> *Id.* at 76.

<sup>59</sup> Hellyer, *supra* note 49, at 298.

Essentially, the citation stickiness metric allows us to examine the work of both practicing attorneys and judges in the adversarial system.<sup>60</sup> And in the context of our question, it allows us to study the level of agreement between each party and the court about what cases are relevant enough to cite when litigating and resolving a dispute.

Interestingly, Berring highlighted the importance of using court opinions and briefs to study the meaning of a court's decision.<sup>61</sup> And the original citation stickiness article concluded that “the variety of research tools and methods” may explain differences in rates of citation stickiness.<sup>62</sup> It seems, then, that our metric might expose data better able to tell us about the structure of legal information, legal research, and the law's development than those used to do so in the past.

### A. About citation stickiness

A citation is “sticky” if it appears in a court opinion and at least one party's brief.<sup>63</sup> Sticky citations show how often a court cites the same authorities as at least one of the litigants.

Endogenous citations are citations that appear for the first time in an opinion, springing from the court itself.<sup>64</sup> These citations, necessarily, are included as a result of independent research by courts.

Super-sticky citations are citations cited in both parties' briefs and then again in the court's opinion.<sup>65</sup> These are cases that all involved—the adversarial parties and the court—think are important to resolving the dispute.

### B. Our methodology

As much as possible, we followed the same methodology as the original citation stickiness study.

For our dataset, we selected Fourth Circuit cases from 1957 for a few reasons. First, and most importantly, we had access to historical Fourth Circuit briefs in our home institution's law library collection. Second, we wanted to be able to compare our data to 2017 data from the original

60 Kevin Bennardo & Alexa Z. Chew, *Citation Stickiness*, 20 J. APP. PRAC. & PROCESS 61, 67 (2019) (“[C]itation stickiness is worth studying because it provides a window into judicial decisionmaking. Judges often lament the quality of attorneys' briefs. Attorneys often lament the quality of judges' decisions, especially when the opinions explaining those decisions veer away from the issues set forth in the briefs.”).

61 Berring, *Search for Cognitive Authority*, *supra* note 6, at 1703–04 (“The typical decision contains the reasoning of a judge or judges, answering problems raised in the briefs of parties on appeal. . . . The considerable work done by appellate attorneys does not travel with the case. Nor do links to the various sources the attorneys used.”).

62 Bennardo & Chew, *supra* note 60, at 108.

63 *Id.* at 64.

64 *Id.*

65 *Id.* at 84.

citation stickiness study. Finally, for the year, we wanted to choose a time definitively in the pre-CALR era.

We began our search for a sample in Westlaw’s cases database. We narrowed to Fourth Circuit cases and then ran a plain language query of “1957” to ensure that all cases had at least that string of numbers within the document. We then filtered the results by date for 01/01/1957–12/31/1957 to ensure our 1957 cases were indeed decided in 1957. This gave us 181 total cases.

Then, we eliminated cases where there would not be a full opinion or where there might be confounding “noise” from briefing by nonparties. So within our results we searched for “curiam OR amicus OR amici” and eliminated any cases returned. There were 74 cases matching, so we subtracted those from our total, leaving us with 107 cases.

We also had to figure out how to get citations from the parties’ briefs reliably. Since briefs from 1957 are not available on Westlaw, we relied on the print collection of Fourth Circuit briefs at the University of North Carolina’s Kathrine R. Everett Law Library. These briefs were conveniently located at our institution and could be scanned on-site for data collection. Local court rules also required that parties create tables of authorities cited and include them with their filings.<sup>66</sup> Like the original study, we excluded cases in which there were supplemental briefs or amicus briefs in order to capture cases progressing along the traditional pathway of appellant brief, appellee brief, and (when included) appellant’s reply brief.<sup>67</sup> One case was also excluded as one of its briefs cited no cases.<sup>68</sup>

We verified that the briefing in each case met our criteria. We also excluded cases from 1957 if all briefs were not available in typeset format in the print collection. The title page of each brief was scanned as was the table of authorities cited. The unique citations from the tables of authorities were entered into our spreadsheets.

To collect the 1987 dataset, we followed the same procedure as for the 1957 dataset except to substitute 1987 for 1957 in the Westlaw searches and filters. Like the 1957 briefs, the 1987 briefs are not on Westlaw but are in our institution’s print collection. An in-depth description of data collection from the 2017 cases can be found in the original citation stickiness publication.<sup>69</sup> The main difference among the dataset collections, however, is that unpublished opinions also needed to be removed from the samples in 1987 and 2017.

---

<sup>66</sup> REVISED RULES OF THE UNITED STATES COURT OF APPEALS FOR THE FOURTH CIRCUIT, Rule 10 §§ 2(a), 4(a), 5 (1952).

<sup>67</sup> Bennardo & Chew, *supra* note 60, at 79.

<sup>68</sup> Brief of Appellee, *United States v. One 1955 Model Ford Convertible Auto.*, 241 F.2d 86 (4th Cir. 1957).

<sup>69</sup> Bennardo & Chew, *supra* note 60, at 78–81.

At first, we planned to select the first 25 cases from each year 1957 and 1987 because the original study used 25 cases from each circuit, including the Fourth. That sampling method was chosen to ensure diversity of subject matter and a practically (but not perfectly) random sample. That said, then as now—the number of citations, not the number of cases—is the relevant sample size. After beginning our data analysis and realizing that each case’s opinion had far fewer citations in 1957 and 1987 than in 2017, we increased the number of cases we reviewed so that the sample size of citations would be closer to the 2017 sample sizes.

Now, for the size of our samples. The 25 cases from 2017 contained 436 unique citations to decisional authority.<sup>70</sup> The briefs in those cases contained 2,002 unique citations to decisional authority. The 28 cases from 1987 contained 236 unique citations to decisional authority. The briefs in those cases contained 1,018 unique citations to decisional authority. The 27 cases from 1957 contained 309 unique citations to decisional authority. The briefs in those cases contained 1,057 unique citations to decisional authority. As in the original citation stickiness study, the relevant sample sizes are the numbers of unique citations in judicial opinions and the number of unique citations in briefs. The sample sizes were large enough to show significant differences in the stickiness rates as measured by confidence intervals.<sup>71</sup>

### III. Results

The results of our citation study surprised us. We hypothesized that we would see a higher rate of citation stickiness in pre-CALR opinions based on the more coherent nature of case-finding done using the digests and with a more limited set of published authorities to draw from. In fact, we found that the opposite was true. The rate of citation stickiness was lower in the earlier cases, and higher in the post-CALR opinions. For the 1987 cases, decided right in the middle of 1957 and 2017, the rate of stickiness was also in the middle.

70 Decisional authorities result from decisions made by judges and similar decisionmakers. See Bennardo & Chew, *supra* note 60, at 81 n.77; see also William H. Manz, *Citations in Supreme Court Opinions and Briefs: A Comparative Study*, 94 LAW LIBR. J. 267, 267–68 (2002). The Manz study included citations to judicial opinions and administrative decisions, but excluded citations to constitutions, statutes, and regulations. Manz, *supra* note 70, at 268.

71 As in the original *Citation Stickiness* article, we calculated 95% confidence intervals using the Exact test in Stata. See Bennardo & Chew, *supra* note 60, at 83. As the original article explained, “A confidence interval expresses the percentage probability that data lies between two limits.” *Id.* at 83 n.80 (citing ALAN R. JONES, PROBABILITY, STATISTICS AND OTHER FRIGHTENING STUFF 102 (2019)).

### A. Some specifics

In the U.S. Court of Appeals for the Fourth Circuit cases sampled from 2017, 55% of citations in the opinion were sticky, meaning they were cited in at least one party's brief. In 1957 cases sampled from the same circuit, only 44% of citations in the opinions were sticky. This is a significant difference.<sup>72</sup> In 1987 cases from the Fourth Circuit, 48% of citations in the opinions were sticky. This is not a significant difference from the 55% stickiness rate in the 2017 cases or the 44% stickiness rate in the 1957 cases.<sup>73</sup>

What this means is that the court in 1957, before the advent of computer-assisted legal research, introduced cases to its opinions without those cases having been raised in either party's brief 56% of the time. In 1987, the court did this 52% of the time. And in 2017, the court did this only 45% of the time. The court, then, was more likely to identify relevant authority on its own—endogenously—when the universe of case finding tools was more unified and the number of available cases was smaller.

When looking at “super sticky” citations, our findings show a similar trend of disagreement over relevant decisional authority and, perhaps, incoherence in legal doctrine in pre-CALR cases when compared with post-CALR cases. In 2017, unique cases cited in court opinions appeared in both parties' briefs 28% of the time. In 1987, unique cases cited in court opinions appeared in both parties' briefs 22% of the time. And in 1957, unique cases cited in court opinions appeared in both parties' briefs only 15% of the time. In other words, nearly 3 out of every 10 cases cited by a court were also cited by both parties in 2017, whereas in 1957, that happened about 3 out of every 20 times—or half as often.

Put yet another way, imagine that, after the attorneys for both sides of a case thoroughly researched and argued their sides to the Fourth Circuit in 1957, both attorneys sat down together to read the court's opinion. Our results show that 8.5 times out of 10, at least one of the attorneys might think, why didn't *I* cite that case?

---

**72** The 95% confidence intervals do not overlap for these two sets of citations:

1957: 38.71%–50.07%  
2017: 50.24%–59.78%

**73** The 95% confidence intervals do overlap for the other pairs of citations:

1957: 38.71%–50.07%  
1987: 41.36%–54.46%  
2017: 50.24%–59.78%

Note that the confidence interval is much tighter for 2017 than for the earlier years; this is a function of the sample size (436 case citations in the opinions) being about 50% larger than the sample size of the earlier years (309 for 1957 and 236 for 1987).

	1957	1987	2017
Unique citations	1,229	1,141	2,198
Sticky citations: appeared in opinion and at least one brief	44%	48%	55%
Super sticky citations: appeared in opinion and both briefs	15%	22%	28%
Citations in briefs that appeared in opinions	13%	11%	12%
Average number of cases cited per opinion	11	9.4	17.4
Average number of cases cited by parties	37.8	40.7	80.1
Average number of sticky cites per opinion	4.9	4.5	9.6
Average number of endogenous cites per opinion	6.1	4.9	7.8

## B. Some interpretations and wild speculation

Our results show how often the Fourth Circuit cited to the same authorities as the parties at three moments in time.<sup>74</sup> Over our sixty-year study period, we observed that stickiness increased from 1957 to 1987 to 2017. This at least means that, pre-CALR, there was less coherence or agreement between advocates and courts than previously believed. This is counter to much of the commentary.

We think this is an interesting finding on its own, but inquiring minds want to know why citation stickiness increased over this period, even though the dominant theory predicted that citation stickiness would decrease as CALR exploded the cozy universe of thinkable thoughts. We have some ideas, which you can read once you finish this paragraph. But first, a few things are probably *not* causing the upward trend.<sup>75</sup> We can probably eliminate some causes based on prior research: Per Fronk, changes in judicial style, workload, and so on are unlikely drivers of citation stickiness.<sup>76</sup> Other unlikely drivers include individual judge characteristics, such as experience, party affiliation, or judicial role, per the original citation stickiness study.<sup>77</sup> Now, on to the causes that have more potential.

First, researchers might be converging on the same cases because tools measuring depth of treatment were easily available in 2017. Hanson argued that a big problem with the digests is that there was “no evaluative

<sup>74</sup> See Bennardo & Chew, *supra* note 60, at 105.

<sup>75</sup> For our thoughts on what could be causing the upward trend, see *infra* section III.C.

<sup>76</sup> Fronk, *supra* note 54, at 79.

<sup>77</sup> *Id.* at 110–11.

component” with the case-finding tool that would “help the researcher separate the important [cases] from the vast majority that merely mentioned the relevant point of law without making a notable contribution to it.”<sup>78</sup> Now, on Westlaw for example, word searches, citators, and tables of authority all produce lists of cases with an icon indicating depth of treatment. The “atleast” connector also makes it easy for researchers to limit results to cases that use a particular word many times, a rough proxy for depth of treatment in the word search context.

Second, stickiness might have risen in the Fourth Circuit over the past 60 years because of the increasing rigor in research and writing instruction during law school.<sup>79</sup> This rigor has become more uniform across law schools in the past 30 years, which might lead to attorneys and judges using similar methods for locating relevant precedent, which leads to similar research results and thus cited cases. For example, more attorneys and judges would have learned how to use depth of treatment tools during law school, both because these tools exist now and because research instruction has increased.

Third, perhaps because of reasons one and two, lawyers might be better now at finding cases that judges agree are relevant enough to include in their written decisions. The average number of sticky cites per opinion doubled from 1957 to 2017, going from 4.9 sticky cites per opinion to 9.6. This increase tracks the increase in the average number of cases that the parties cite, which has also doubled from 1957 to 2017, going from 37.8 cases to 80.1. However, the average number of citations in opinions did not increase at the same rate: the 2017 opinions had about 1.6 times the number of citations as the 1957 cases. So, by doubling the number of cases cited in briefs, parties have doubled the number of sticky cases in those briefs, even though the percentage of sticky cases cited in the briefs has stayed the same.

Fourth, Fronk’s documented decrease in string cite usage as a percentage of overall opinion cites could increase stickiness by limiting the number of new cases that a court introduces by string cite.<sup>80</sup> Fronk also reasoned that the increase in expository citation suggested that judges were spending more research energy per cite, despite a “caseload explosion” of 630% from 1955 to 2005.<sup>81</sup> If Fronk is correct that the

78 Hanson, *supra* note 14, at 569.

79 The Fourth Circuit does not track exactly with the results of Marvell’s citation stickiness study from the early 1970s, which found a citation stickiness rate of 55% for 30 Sixth Circuit civil opinions issued in 1971 and 1972. See THOMAS B. MARVELL, APPELLATE COURTS AND LAWYERS: INFORMATION GATHERING IN THE ADVERSARIAL SYSTEM 134–36 (1978).

80 See Fronk, *supra* note 54, at 69.

81 *Id.* at 79.

research cost per cite has increased over time, then the cost of finding and adding endogenous citations to an opinion would likely be higher than the cost of adding sticky citations that have already been vetted and described by the parties. However, this explanation seems less convincing given the increase in number of endogenous cites in opinions has increased (although not by much) from 6.1 in 1957 to 7.8 in 2017.

Fifth, it is also possible that CALR algorithms have changed to push advocates and courts closer to one another. Having adjusted to the wilds of text searching, research scholarship has turned to how research platforms' algorithms influence research results. While search algorithms rank results differently across platforms,<sup>82</sup> perhaps *within* a platform, like Westlaw, the results of case law searches are more uniform than the results generated using a print digest.<sup>83</sup> In simple terms, attorneys and judges in 2017 might have been seeing more of the same cases in their research of a topic than technology had allowed before, by virtue of improved (or at least more consistent) search algorithms on the same platform. This is a fertile and growing area of scholarship in legal information, and more study is needed to determine the degree of algorithmic influence on legal citation practices.

Finally, judges in 2017 might have been purposefully limiting endogenous citations as part of an overall trend towards judicial minimalism. Judicial minimalism, nicely summarized by Lauren Cyphers in her student note, "is a case-by-case approach that looks only to the specific set of facts before it and crafts a decision narrowly tailored to those unique facts."<sup>84</sup> Delgado and Stefancic raised this possibility in 2007, concluding that electronic searching can "lead to judicial minimalism—narrow, fact-based decision making that ignores emerging legal theories and decides cases on the narrowest possible grounds."<sup>85</sup> Their reasoning was that CALR was better at finding concrete examples than abstract patterns, and thus fact-based searching "can easily cause you to miss [a new legal theory] that is emerging in another jurisdiction."<sup>86</sup> This reasoning aligns

82 Susan Nevelow Mart, *The Algorithm as a Human Artifact: Implications for Legal [Re]Search*, 109 LAW LIBR. J. 387 (2017).

83 However, research into the use of natural language processing posits that algorithms using the technology "ground [themselves] in the forms and functions of cognitive authority of the past—perhaps such as giving cognizance to most-cited cases, adhering to jurisdictions, performing citation analysis, building on West's Topic and Key Number System, emphasizing cases annotated in *American Law Reports*, or any number of a hundred factors that make up the current terrain of the legal information environment." Paul D. Callister, *Law, Artificial Intelligence, and Natural Language Processing: A Funny Thing Happened on the Way to My Search Results*, 112 LAW LIBR. J. 161, 167 (2020).

84 Lauren Cyphers, Note, *Maximalist Decision Making: When Maximalism Is Appropriate for Appellate Courts*, 123 W. VA. L. REV. 611, 612 (2020) (citing CASS R. SUNSTEIN, *ONE CASE AT A TIME* ix–x (1999)); Neil S. Siegel, *A Theory in Search of a Court, and Itself: Judicial Minimalism at the Supreme Court Bar*, 103 MICH. L. REV. 1951, 1952 (2005)).

85 Delgado & Stefancic, *supra* note 20, at 323–24 (suggesting Margaret J. Radin & Frank Michelman, *Pragmatist and Post-structural Critical Legal Practice*, 139 U. PA. L. REV. 1019 (1991), to learn more about the philosophy of legal minimalism).

86 *Id.* at 324.

with Krieger and Kuh’s study results from a few years later, that electronic researchers used fact-based searching far more than paper researchers.

### C. Wild speculation about creating a coherence measurement

It appears that we cannot yet use our results to show that stickiness means a shared sense of relevance, that is, whether the parties and court share a cozy universe of legal authorities. Yet we also wonder why stickiness wouldn’t mean exactly that? Wouldn’t a focus on appellate matters studied show more of the “legal concepts” than the facts?

For example, imagine a new metric called “party stickiness” or “party coherence” that looked at the number of times both parties cited a case, and whether the court also cited it. The number of times both parties cited a case would be the numerator of our new metric, but the denominator could be several things: the number of unique cases cited in the briefs, the number of unique cases cited in the briefs and opinion, or even the number of unique cases cited in the opinions. We did those calculations with our data, but we are still thinking about what they might mean, if anything. They are in the table below.

	All	1957	1987	2017
Cases cited by both parties	637	130	169	338
Cases cited by both parties and opinion	222	45	51	126
Cases cited by both parties but not the opinion	415	85	118	212
Percent of cases cited by both parties out of all brief cites	16%	12%	17%	17%
Percent of cases cited by both parties but not in the opinion	10%	8%	12%	11%
Percent of cases cited by both parties compared to number of opinion cites	65%	42%	72%	78%

To see whether citation stickiness could measure coherence will likely require looking at how courts *use* the sticky citations in their opinions, not just counting them. Both this study and the original citation stickiness study shied away from studying use because it is so time consuming. However, Brian N. Larson set out to do just that in two ambitious papers: *Precedent as Rational Persuasion*<sup>87</sup> and *Endogenous & Dangerous*.<sup>88</sup> Larson’s studies analyzed federal district court opinions addressing

<sup>87</sup> Brian N. Larson, *Precedent as Rational Persuasion*, 25 LEGAL WRITING 135 (2021).

<sup>88</sup> Brian N. Larson, *Endogenous & Dangerous*, 22 NEV. L.J. \_\_\_\_ (forthcoming 2022).

dispositive motions, with the specific legal issue studied being the affirmative defense of fair use to copyright infringement.<sup>89</sup> Although these first two studies used federal district court opinions, Larson is following these matters through appeal (if any), and the results of that stage of his study could help refine a coherence measurement, particularly because Larson expressly engages with the literature on citation stickiness and endogeneity.<sup>90</sup>

#### **D. So many ideas for future study**

This article asks a narrow question and does its best to answer that narrow question. However, it has generated many other questions that might be answerable with our dataset.

We describe some of those future research questions below and intend to broaden the scope of our project to address them in a longer article. For the richness of these questions, we are particularly grateful to the participants of the Little Boulder Conference with whom we workshopped this paper, to Brian Larson and the faculties at Texas A&M University School of Law and Drexel University Thomas R. Kline School of Law who workshopped this paper with us, and to the attendees and organizers of the Yale Virtual Symposium on Citation and the Law.

##### **1. Increase our sample size for 1957 and 1987**

The sample sizes for this symposium paper are uneven, and we would like to increase the sample sizes for 1957 and 1987 to be closer to the sample size for 2017. Doing so would require adding about 1,000 more opinion citations for each of 1957 and 1987. Gathering the opinion citations itself is not that difficult because they are available on Westlaw. But gathering the parties' citations for each of those opinions must be done by hand, using paper copies of the briefs that are archived at our university. Although the stickiness percentages for 1957 and 2017 are significant when considering 95% confidence intervals, larger sample sizes should improve the precision of our stickiness calculations.

##### **2. Figure out that coherence measurement**

We recognize that our analysis of a stickiness-based coherence measurement is incomplete. With some more thinking, we hope to complete the analysis and identify a useful measure of coherence.

.....  
<sup>89</sup> *Id.* at \_\_\_\_.

<sup>90</sup> Another recent citation study does analyze how judges use citations in their opinions, but it does not engage with either Larson's work or Bennardo & Chew's. See Mark Cooney, *What Judges Cite: A Study of Three Appellate Courts*, 50 STETSON L. REV. 1 (2020).

### 3. Test whether endogenous cites support procedural rules

Since the first presentation on the initial data of the first citation stickiness paper, the most-asked question is how many of the endogenous cites are cases that support procedural rules like the standard of review. Because attorneys and chambers can have stock language that they use to describe procedural rules, one set of cases used to describe the 12(b)(6) standard could be entirely different from another set of cases used to describe the same standard in basically the same way. Differences in stock procedural language would lead to lower stickiness without a difference in meaning, or even meaningful research.

Since the original stickiness paper was published, Brian Larson has created a coding system for categorizing how courts use each statement of law and citation in a judicial opinion. We can use Larson's system to code our Fourth Circuit data set to look for procedural rules and their uses. This would help answer the most frequently asked question and tie our study more closely with Larson's ongoing study of endogenous citations.

With this later analysis in mind, we did a small pilot study using the ten opinions with the most endogenous citations in them from 1957, 1987, and 2017. These opinions yielded 128 endogenous citations in 1957, 95 in 1987, and 136 in 2017. We asked our research assistant to go through those thirty opinions and identify endogenous citations that obviously supported an appellate standard of review. We asked him to look for the "obvious" ones because sometimes reasonable minds can disagree as to whether a statement of law is "procedural" or "substantive." The results of this informal pilot showed an increase in endogenous procedural citations over time: 3% of endogenous cites in 1957, 14% in 1987, and 18% in 2017.

The 95% confidence interval for the 2017 percentage is 12.3%–25.9%, which suggests that procedural "boilerplate" accounted for a chunk of the endogenous cites in the original 2017 study. By contrast, the confidence interval for the 1957 percentage dips down to nearly zero. Because so few endogenous cites in 1957 were procedural, these initial results suggest that the increase in stickiness is not related to procedural citations. If nothing else, this pilot suggests a notable change in the way courts cite cases to support the standard of review from 1957 to 2017.

### 4. Analyze the weight of the endogenous authorities

Another frequently asked question is what courts the endogenous citations come from. This question is also one that is commonly addressed in citation studies but that neither this study nor the original citation stickiness study sought to answer. Answering this question for our data set would again both sate the curious minds of our audience and also tie our study in with other citation studies, particularly Larson's.

Our small pilot study described just above included identifying each endogenous citation's issuing court. Our initial results are summarized in the table below and show an increase in endogenous cites to other Fourth Circuit cases and a decrease to other circuit cases. The other court categories don't suggest a pattern.

<b>Cited Court</b>	<b>1957</b>	<b>1987</b>	<b>2017</b>
U.S. Supreme Court	20%	34%	16%
Fourth Circuit	10%	22%	36%
Other federal circuit	33%	26%	10%
Federal district court	12%	12%	15%
State high court	15%	4%	13%
State intermediate appellate court	1%	0%	4%

##### **5. Analyze the frequency of endogenous citations in string citations**

Given Fronk's findings, string citations could be a large source of endogenous citations in judicial samples. His study suggests that the percentage of opinion cites that exist only in string cites would be highest in the 1957 cases, much lower in the 1987 cases, and lower still in the 2017 cases. That Fronk's study also used years ending in seven is particularly fortuitous for comparing his results and ours.

Our small pilot study included this string cite analysis. Our initial results track Fronk's findings and are summarized in the table below. In addition to string citations, we counted endogenous citations that appeared only as citing or quoting parentheticals or only as part of a quotation. One observation is that the 1957 opinions included 21 endogenous citations in footnotes, which decreased to 9 in 1987 and only 1 in 2017. Some of these footnoted citations were also string citations.

	<b>1957</b>	<b>1987</b>	<b>2017</b>
Only in a string citation	58%	49%	21%
Only in a citing parenthetical	0%	1%	4%
Only in a quoting parenthetical	0%	1%	7%
Only in a quotation	8%	2%	0%
Only in subsequent history	< 1%	0%	< 1%
Only in footnote	16%	9%	< 1%

## 6. Analyze Judge Widener’s use of endogenous citations

Fronk’s study analyzed the individual citation practices of two long-serving circuit court judges, including one from the Fourth Circuit, Judge H. Emory Widener Jr.<sup>91</sup> Judge Widener served on the Fourth Circuit from 1972 to 2008, and Fronk analyzed 397 of Judge Widener’s majority opinions, which had 4,393 unique case citations.<sup>92</sup> Fronk found that Judge Widener’s citation patterns changed across time, closely matching the aggregate data that Fronk collected.<sup>93</sup> For example, his use of string citation steadily declined from over 21 percent in 1972–1977 to under 9 percent three decades later.<sup>94</sup> One of Fronk’s takeaways from his longitudinal looks at two judges’ citation practices is that CALR might have had a “conforming” effect on judges’ citation practices.<sup>95</sup>

Again, because our dataset overlaps with Fronk’s, we could add on to his longitudinal study of Judge Widener’s citation patterns by calculating the stickiness of his opinions over that same time period. This might tell us something about the connection between the changes in judicial citation practices that Fronk observed and courts’ independent research.

## 7. Look at historical research instruction practices

One potential reason that citation stickiness has increased over time is a change in legal research instruction to be more uniform. And with respect to coherence, more uniform research instruction seems more likely to result in greater coherence. This study did not look at historical research instruction practices to see if they match that theory, but a future study could.

## 8. Look at historical court rules for citation

Current federal court rules require parties to substantiate their arguments with citations to relevant legal authorities. But Fronk’s study shows that the ways judges cited legal authorities changed across time. Studying historical court rules could lend insight into the ways that *parties* cite legal authorities.

## 9. Study opinions with novel legal theories

A recurring concern with both print-era research methods and CALR is that they stifle innovative legal theories and, specifically, innovative jurisprudence.<sup>96</sup> If so, judicial opinions that advance novel legal theories

91 Fronk, *supra* note 54, at 80.

92 *Id.*

93 *Id.* at 84.

94 *Id.*

95 *Id.* at 87.

96 See generally Nicholas Mignanelli, *Critical Legal Research: Who Needs It?*, 112 L. LIBR. J. 327 (2020).

should have more endogenous citations. But if innovative jurisprudence grows from advocates' efforts, the judicial opinions should have fewer endogenous citations. A future study could, then, focus on opinions that advance novel legal theories, perhaps as noted by legal scholars, including student notes and recent developments. This would show whether the cites that advance that novel theory are sticky or endogenous.

#### **10. Repeat the study in 2037**

By the time we finish with future studies 1 through 9 above, it will probably be time to add another 20-year block to our study!

### **Conclusion**

In this study, we sought to bring together several strands of legal scholarship: theory about the effect of CALR on legal research, studies of research and citation practice by courts, studies of research practice by attorneys and law students, and studies directly comparing court citations and party citations in the same matter. Our primary empirical question was straightforward: during the print era, did courts cite the same cases as the parties more often than during the digital era, as posited by the universe of thinkable thoughts theory? The answer was similarly straightforward: No. The results show that, pre-CALR, there was less agreement between advocates and courts than previously believed by many commentators. If a limited universe of thinkable thoughts existed in the print era, it was not cozy enough for the attorneys and judges to cite the same cases during the appellate process.