USING TEXT AS DATA TO MEASURE LATENT LEGAL CONSTRUCTS: A DICTIONARY-BASED APPROACH

Justin Wedeking[*] *& Alexander Denison*[**]

2017 MICH. ST. L. REV. 1057

TABLE OF CONTENTS

INTRODUCTION

In recent years, two large movements have been making distinct contributions to the identity of the legal world. These movements, empirical legal studies[1] and legal analytics, have done much to advance our study and understanding of the law, as well as our ability to predict future legal events of interest. It is true that these movements may have occasionally competing or contradictory goals (e.g., prediction versus explanation), but for the most part, it is their shared commonality that is driving both of them forward. This commonality is a strong emphasis on systematic and rigorous methods that emboldens scholars with a new confidence for answering their legal questions. While it may be near impossible in the present space to cover completely the points and nuances of each movement along with their differences and similarities, we think it is

---

    *   Associate Professor, Department of Political Science, University of Kentucky.
    **   Ph.D. Candidate, Department of Political Science, University of Kentucky.
    1.   Theodore Eisenberg, *The Origins, Nature, and Promise of Empirical Legal Studies and a Response to Concerns*, 2011 U. ILL L. REV. 1713, 1713-14.

beneficial to highlight one technique that is of interest and value to scholars of both movements.

In this Article, we showcase a measurement approach that uses text as data to create a scale, or what is sometimes referred to as estimating a latent construct. What is different about our approach is that it is premised on using word dictionaries to define the two ends of the spectrum. We offer it as a simple yet powerful alternative. Ironically, while there are exceptions,[2] dictionary-based approaches are not something that measurement scholars generally consider when trying to create scales.[3] We apply our measurement technique to an important substantive question involving news media coverage of Supreme Court decisions, though we note at the outset that our technique could be applied more broadly. Specifically, we ask whether news content of the U.S. Supreme Court is ideologically slanted or biased. This is an important question because prior research has shown that ideological bias is prevalent in other areas of political news coverage,[4] yet because the Court plays a special role in our government in terms of the rule of law, a biased news media covering Court decisions strikes at the foundational underpinnings of democracy. At the same time, we also argue there may be reasons why reporting on the Supreme Court may be different in this regard.

To answer our question, we use two different dictionaries—one based on "ideological" words and another based on "partisan" words—to examine the content of internet news stories from twenty-six different outlets that covered decisions from the 2014 Supreme Court term. Next, we generate a series of plausible estimates based on different text processing decisions and then compare our estimates to externally valid measures of the ideology of news outlets.

Our contribution is important for several reasons. First, measuring latent constructs is a necessity for both empirical legal studies and legal analytics. Indeed, measurement is an important fundamental task for any empirical inquiry and is essentially the first step before any analysis can be undertaken. For example, before

---

2. *See generally* Michael Laver & John Garry, *Estimating Policy Positions from Political Texts*, 44 AM. J. POL. SCI. 619 (2000).

3. Justin Grimmer & Brandon M. Stewart, *Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts*, 21 POL. ANALYSIS 267, 268-69 (2013).

4. Matthew Gentzkow, Jesse M. Shapiro & Daniel F. Stone, *Media Bias in the Marketplace: Theory* (Nat'l Bureau of Econ. Research, Working Paper No. 19880).

answering whether and why legal expertise among judges has changed over time, one first has to be able to measure legal expertise in a way that is dynamic over time.[5] Second, the technique we use here could be easily extended to measure any other latent construct of interest. The approach provides a very simple and easy-to-use framework, with its key being the identification of words that occupy both ends of the spectrum. It makes no assumptions about the ideological leanings of the *documents* in the text, avoiding one of the assumptions made by a classical supervised learning approach. Yet, importantly, because the dictionaries bring information to the process itself, it is not a completely unsupervised approach, where the researcher has no assurance that what one is measuring is in fact the dimension of interest. Ultimately, we are able to assess the validity of the approach with external measures of interest and are able to provide an answer to our question—whether news content of the Court is slanted or biased.

I. MEDIA COVERAGE OF JUDICIAL EVENTS

An underdeveloped subject area in empirical legal studies deals with the news media's treatment of judicial activities and, more pertinent to our mission, the potential ideological biases that may seep into coverage of judicial institutions like the U.S. Supreme Court. Extant studies reveal Supreme Court news coverage to be largely inadequate[6] and susceptible to conflict-centric biases in reporting.[7] While coverage of the Court may be somewhat different than that of other governmental activities due to the relative inaccessibility of the institution and its output,[8] the importance of understanding how the media covers judicial events is nonetheless vital to understanding public sentiment toward the Court. First, most Americans receive all of their information about the Supreme

---

5. Anne Lippert & Justin Wedeking, *Is Judicial Expertise Dynamic? Judicial Expertise, Complex Networks, and Legal Policy*, 2016 MICH. ST. L. REV. 567, 573.

6. Chester A. Newland, *Press Coverage of the United States Supreme Court*, 17 WESTERN POL. Q. 15, 16 (1964).

7. MICHAEL A. ZILIS, THE LIMITS OF LEGITIMACY: DISSENTING OPINIONS, MEDIA COVERAGE, AND PUBLIC RESPONSES TO SUPREME COURT DECISIONS 25 (Univ. of Mich. Press, 2015); Tyler Johnson, *How and Why the Supreme Court Remains Undercovered*, *in* COVERING THE UNITED STATES SUPREME COURT IN THE DIGITAL AGE 23-24 (Richard Davis ed., 2014).

8. *See generally* RICHARD DAVIS, DECISIONS AND IMAGES: THE SUPREME COURT AND THE PRESS (Englewood Cliffs: Prentice Hall 1994).

Court's activities from the media environment that not only chooses what events or decisions to spend resources covering, but how they opt to frame their content as well.[9] As opposed to reading Supreme Court opinions directly, most rely on the media to interpret the decisions and supply enough adequate information for knowledgeable assessments of the institution and its actions. Thus, the media's role in providing access to the activities of the Supreme Court is a tremendously important piece of the relationship between the Court and public support, both specific and diffuse.

Second, the ways in which these media outlets opt to present their information, through potential subjectivity in coverage, bias, or slant, can have major impacts on public perceptions and, subsequently, judicial behavior. For instance, Bartels and Johnston[10] find that individuals who receive the majority of their news from "sensational" sources like cable news and talk radio are less supportive of the Supreme Court than those who prefer more traditional news sources like newspapers and network news programming. This theoretical difference between sensational and traditional news, according to the authors, is a matter of tone. And with the information provided by sensational sources likely to be more negative in tone, it thus influences consumers to be more critical of institutions like the Court. While Bartels and Johnston do not go so far as to validate the negativity of sensational media sources, others have used computer-assisted text analysis to find that negativity in tone varies systematically. For example, Denison, Wedeking, and Zilis find negative tone to be a function of the negativity of Court output, the degree of *dissensus*, and whether coalitions on the Court fall along ideological lines.[11]

Negativity is, however, only one latent construct through which we can observe differences in media coverage of the Supreme Court and is not necessarily an ideal proxy for critical or partial coverage. Perhaps of greater concern is the extent to which media coverage of

---

9.   As we discuss at the end of the Article, we only focus on content slant or bias in this Article, and we leave for future research how coverage slant or bias may also contribute to media bias.

10.   Christopher D. Johnston & Brandon L. Bartels, *Sensationalism and Sobriety: Differential Media Exposure and Attitudes Toward American Courts*, 74 PUB. OPINION Q. 260, 261 (2010).

11.   Alexander Denison, Justin Wedeking & Michael A. Zilis, *Negative Media Coverage of the Supreme Court: The Interactive Role of Opinion Language, Coalition Size, and Ideological Signals* (July 30, 2016), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2816431 [https://perma.cc/7P3W-3YUH].

the institution is subject to ideological slant. Recent research from Ho and Quinn suggests that media coverage of the Supreme Court is indeed subject to systematic biases, most directly linked to the media outlet's ideological affiliation.[12] The authors utilize an item response theory (IRT) framework that places the presence and position of editorial content in several major newspapers against the votes of Supreme Court justices themselves, finding media outlets perceived to be liberal in content most closely aligned with liberal justices on the Court, and vice versa.[13] For instance, *The New York Times*, considered to be on the left of the ideological spectrum, was most closely aligned with the voting record of John Paul Stevens, a justice widely considered to be a liberal on the Court.[14] Similarly, *The Wall Street Journal*, an outlet often associated with conservative content, was most closely aligned with the record of Justice Scalia, a longtime conservative stalwart.[15] This provides one of the more robust analyses of media slant, not only in editorial coverage of the Supreme Court, but the broader media environment as well.

While Ho and Quinn's work is certainly a novel approach to measuring editorial bias, the assumptions underlying this finding, among similar findings in the literature regarding media coverage, raise questions about one of the key inferences that was drawn.[16] Of paramount concern is the use of *editorial* content as opposed to that of "straight" news content in the analysis. The distinction between the two is important. Editorials are largely written to be persuasive in tone, and the ideological bias found within many editorials should come as no surprise to most. Indeed, readers expect editorials to have a particular slant on any given issue. A column from George Will in *The Washington Post* will often take the ideologically conservative position on an issue and advocate accordingly, just as an article from Eugene Robinson in the same paper will advocate for the liberal position under most circumstances. This content is essentially walled off from what is generally considered the "news" aspect of the outlet, usually put on a page of its own or in a separate and clearly distinct section. Alternatively, what is "straight" news—in our examples later, the general reporting on Supreme Court decisions—is arguably less likely to be so ideological in tone, at least if one believes in the

---

12.   *See generally* Daniel E. Ho & Kevin M. Quinn, *Measuring Explicit Political Positions of Media*, 3 Q.J. POL. SCI. 353 (2008).
13.   *See id.* at 356.
14.   *See id.* at 363-64, 370-71.
15.   *See id.*
16.   *See id.*

mission of journalism and the journalistic norms that are espoused by those in the profession.

Ho and Quinn contend that the wall between news and editorial content is crumbling and provide some correlations between the editorials and external measures of news bias as a point of validation.[17] Unfortunately, these external measures, which are based on public evaluations of media slant in various outlets, do not make it clear what the public is thinking about (editorial news or straight news or some combination of both) when making these assessments. The authors use mondotimes.com user ratings of media organizations on a 1-5, liberal–conservative scale.[18] The concern here is that it is impossible to determine whether individuals are basing their assessments on news content or editorial content or some mix of both. In other words, when an individual is asked to consider whether *The Wall Street Journal* is a liberal or conservative outlet, it is difficult to argue that their perception will ultimately compartmentalize the news and editorial sections of the publication. While the news aspect of the publication may be deemed liberal in tone by the respondent, the editorial aspect may be overwhelmingly conservative, and we are unable to properly ascertain what served as the basis for their response. The use of mondotimes.com ratings is not problematic in itself, as we use those ratings below. What is problematic is using them in conjunction with *only editorial* content to then conclude there is a *news content* bias, as opposed to an editorial slant or bias. Thus, while we concede that Ho and Quinn provide a very strong measure of *editorial* bias, it falls short of being a strong measure of *news* bias.[19]

Whether ideological news bias or slant exists, and to what extent, is a matter of primary interest for not only many in the scholarly community but many in the general public as well. Much of this fascination is generated by the supposed role of the free press in American political discourse. As the so-called fourth estate, the media takes on a watchdog function that should be critical of those in power, regardless of the target's ideological persuasion. And yet, prior studies cast doubt on the ability of journalists to put their own ideological preferences aside, particularly when considering the

---

17.   *See id.* at 355.
18.   *See, e.g.*, *id.* at 369-70.
19.   *See id.* at 372-73 (noting several limitations to the study).

overwhelming number of liberal individuals in the profession.[20] As Groseclose and Milyo note, the reportedly disproportionate number of self-identified liberals in the news media makes the "fourth estate" the least representative of any found in the area of governance.[21] Furthermore, political developments since the 2016 presidential election have cast an even greater shadow over the journalistic profession than usual. The Trump Administration's affinity for the phrase "fake news"—frequently used to challenge negative reports of administration activities—has helped to drive down trust in the media and ramp up perceptions of media bias, particularly among self-identified Republicans.[22] As Americans become more ideologically selective in their media consumption,[23] the likelihood of trusting information from ideologically incongruent sources plummets, and the disconnect between Americans on each end of the spectrum may spur greater incivility and polarization.[24]

Measuring news bias is in itself a difficult and controversial endeavor. Since the concept of "bias" is not an observable variable, researchers are forced to rely on the construction of latent variable constructs generated from other observable aspects of media coverage. This is, of course, a common issue in many aspects of the social sciences, particularly when attempting to measure individual attitudes. Mentioned earlier, the mondotimes.com score used by Ho and Quinn offers a simple example of a latent variable used to measure the presence of bias, in which respondents offered their own perceptions of bias in media outlets.[25] While the respondent possesses a true attitude about any individual outlet's slant (*T*), anything from mood to website design to random error (ε) may skew the measurement of *T*. Therefore, what the respondent provides as a score on the website may not, and likely will not, be *T*, but rather the true score influenced by ε. The resulting observed score (*X*) collected

20. *See, e.g.*, ELAINE S. POVICH, PARTNERS & ADVERSARIES: THE CONTENTIOUS CONNECTION BETWEEN CONGRESS & THE MEDIA 140 (1996).

21. *See generally* Tim Groseclose & Jeffrey Milyo, *A Measure of Media Bias*, 120 Q.J. ECON. 1191 (2003).

22. *See* Michael Barthel & Amy Mitchell, *Americans' Attitudes About the News Media Deeply Divided Along Partisan Lines*, PEW RES. CTR. (May 10, 2017), http://www.journalism.org/2017/05/10/americans-attitudes-about-the-news-media-deeply-divided-along-partisan-lines/ [https://perma.cc/QYK8-47QQ].

23. *See generally* Shanto Iyengar & Kyu S. Hahn, *Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use*, 59 J. COMM. 19, 19 (2009).

24. MATTHEW LEVENDUSKY, HOW PARTISAN MEDIA POLARIZE AMERICA 5 (2013).

25. *See* Ho & Quinn, *supra* note 12, at 369-70.

is therefore not necessarily the true score but is believed to be heavily influenced by the true score with some potential variance remaining, as evinced by a simple equation:

$$X = T + \varepsilon$$

None of this is to suggest that user ratings or public attitudes in any capacity are the most ideal way to measure news bias. Again, these are merely the perceptions of the individual, and as mentioned earlier, perceptions of media bias are highly susceptible to individual ideological preferences and are unlikely to be purely objective ratings.[26] Therefore, the development of latent bias constructs—whether from a single observed score or multiple items—leads to a great deal of creativity and controversy. For example, Ho and Quinn pit their findings against two other studies of media bias worthy of note.[27] Peake focuses on the headline tone of front-page newspaper coverage of the Bush presidency, finding that tone is heavily influenced by the ideological position of the paper.[28] Ideological position, in this instance, is based on whether or not the paper endorsed Bush in the election. Unfortunately, the measure of tone, again, does not rely on news content itself but rather a categorical decision made by the individual coder as to whether the *headline* is positive, negative, or neutral based on prior established coding standards.

An alternative bias study from Groseclose and Milyo relies on an examination of think tank citations in congressional speeches, a proxy for the median American, against citations of these think tanks in news reporting.[29] News outlets that reference ideological think tanks to a greater degree than the congressional average are considered to be expressing bias. Groseclose and Milyo rely on a single heuristic without paying attention to actual news content but claim their results are indicative of widespread liberal bias in media coverage.[30] However, the reliance on think tank citations is perhaps doubly problematic because it essentially ignores major aspects of the content itself, and their inclusion in any reporting is not

---

26. *See* Barthel & Mitchell, *supra* note 22.
27. *See* Ho & Quinn, *supra* note 12, at 356.
28. Jeffrey S. Peake, *Presidents and Front-Page News: How America's Newspapers Cover the Bush Administration*, 12 HARV. NAT'L J. PRESS/POL. 52, 52 (2007).
29. *See* Groseclose & Milyo, *supra* note 21, at 1191.
30. *See id.*

necessarily an indicator of agreement but may in fact be included critically.

While we admit that the development of a satisfactory measurement of bias is difficult, and we only scratch the surface in terms of bias studies, we believe there are two essential elements required of a measure of news bias in reporting. First, we believe that a measure of news bias should be based on *news*, not editorial content. It is yet to be clearly demonstrated that editorial bias is commensurate with news bias, and such a relationship should not be assumed. Second, we believe that a measure of news bias should be based on actual news *content*. While prior studies creatively find ways to assess bias through other means, we contend that it is the complete information provided in the news report itself that should be used to determine slant. Yet, given that, it is entirely possible that news coverage of Court decisions may not contain the same ideological slant or bias.

II. WHY MIGHT MEDIA CONTENT ABOUT THE COURT NOT BE IDEOLOGICALLY SLANTED?

There are a number of reasons to consider why Supreme Court content might be comparatively impervious to broad biases. First, despite all of the recent concern about "fake news," mainstream outlets have journalistic norms that are comparable and likely sympathetic to the professional norms of the judiciary. The vast majority of journalists, at least normatively, tend to be concerned with objective truth and public accountability—the "watchdog" function of journalism that is comparable to the counter-majoritarian status of the courts.[31] Also, while many outlets have moved away from the practice of maintaining full-time Supreme Court reporters, those that are in these positions are often trained in law themselves. It is also worth noting that most major journalism programs in the United States require at least some basic coursework in First Amendment or communications law, not only for the purpose of investigative reporting, but for a better understanding of their constitutionally afforded protections.[32] That is to say, many programs

---

31. *See* BILL KOVACH & TOM ROSENSTIEL, THE ELEMENTS OF JOURNALISM: WHAT NEWSPEOPLE SHOULD KNOW AND THE PUBLIC SHOULD EXPECT 140-41 (2007).

32. *See generally* Dianne Lynch, *Above and Beyond: Looking at the Future of Journalism Education*, KNIGHT FOUND., Feb. 19, 2015, https://knightfoundation.org/reports/above-and-beyond-looking-future-journalism-educati [https://perma.cc/Q32Y-CNHM].

feature an emphasis on the history of major legal decisions to familiarize future reporters of standing precedent regarding journalistic controversies. From *Near v. Minnesota* to *Hustler Magazine, Inc. v. Falwell* and beyond, major journalism programs expect students to be able to identify the relevant case law that looms over their profession.

Second, many argue that court reporting, and specifically Supreme Court reporting, is a rather formulaic exercise in which even the most seasoned Supreme Court reporters maintain a basic structure of reporting that rarely deviates from a step-by-step procedure.[33] For example, one reporter said reporting is "almost a formula story: the Supreme Court upheld, struck down, did x, y, z. Then I try to give a sense of the vote and perhaps implications. . . . It's not real hard to do. It's almost like a science."[34] In another example, David Savage, someone who has covered the Court for several outlets, makes the case that *how* the Court is covered suggests a framework that helps resist outside influence. For example, in Davis's edited volume, Savage provides a discussion of how traditional journalists cover the Supreme Court in a new media age. Savage writes, "The basics of writing about the Court have not changed. With each decision comes at least three questions: What did the Court decide? What is the legal basis for the decision? What does it mean?"[35] Building on that last quote, the remoteness of the institution creates a number of obstacles that simply are not found in other avenues of political reporting; thus, the reporter is often left with the task of simply figuring out what the Court actually decided. Like in any other journalistic mission, a certain level of "news judgment" is involved that may lead to ideological discussion, but to a far more limited degree. Finally, with the limited amount of resources dedicated to Supreme Court reporting,[36] the mere lack of space to maneuver into more ideological discussion is an additional constraint. With the exception of some major newspapers, most

---

33. Linda Greenhouse, *Telling the Court's Story: Justice and Journalism at the Supreme Court*, 105 YALE L.J. 1537, 1543 (1996).

34. RICHARD DAVIS, DECISIONS AND IMAGES: THE SUPREME COURT AND THE PRESS 83 (1994).

35. DAVID G. SAVAGE, COVERING THE UNITED STATES SUPREME COURT IN THE DIGITAL AGE 176 (Richard Davis ed., 2014).

36. DAVIS, *supra* note 34, at 83; ELLIOT E. SLOTNICK & JENNIFER A. SEGAL, TELEVISION NEWS AND THE SUPREME COURT: ALL THE NEWS THAT'S FIT TO AIR? 10 (Cambridge Univ. Press 1998); Tyler Johnson, *How and Why the Supreme Court Remains Undercovered*, *in* COVERING THE UNITED STATES SUPREME COURT IN THE DIGITAL AGE 25 (Richard Davis ed., 2014).

outlets do not consistently cover Court activity, and even when they do, the space limitations are considerable.[37]

There is a possibility that certain circumstances may elicit more ideological coverage. It should come as no surprise to even the most casual news consumer that a great deal of news content is based on sensationalistic events of high novelty. Again, in the case of the Court, it is difficult to suggest its activities commonly meet a definition of *sensational*, with the discussion over Scalia's use of the term "jiggery pokery" in *King v. Burwell* standing as a recent high water mark.[38] Alternatively, that third aspect of Savage's to-do list, "what does it mean," opens up some broader avenues for potential ideological content. It is at this point that a story may move from factual analysis to reaction—a point early scholarly research of Supreme Court coverage noted was a particularly fruitful way for journalists to inject some life into Supreme Court coverage.[39]

Clearly, not all cases are likely to warrant a consideration of reaction, just as most cases fail to receive any coverage whatsoever.[40] Regarding this latter point, some research shows there may be a coverage bias in relation to civil rights and liberties cases,[41] suggesting these issue areas attract greater attention from the media, but evidence of any content bias regarding these types of cases is lacking. Instead, it is far more likely that the introduction of ideological content in Supreme Court reporting comes in cases that are conflictual, controversial, or highly salient. For instance, Zilis finds that decisions with low levels of judicial consensus tend to elicit stories that focus in greater detail on dissents, amplifying what is often an ideological divide or conflict on the Court.[42] It is possible that controversial and salient decisions follow a similar trend for a number of reasons. The Court is certainly not blind to the controversy and importance of some of their looming decisions, and content of sharply worded opinions is likely to find its way into news content.[43] Additionally, controversial cases allow for the incorporation of more outside information in a story that may give disproportionate weight to one side of the case. Finally, and in a

---

37. *See* Greenhouse, *supra* note 33, at 1550.
38. 135 S. Ct. 2480, 2500 (2015) (Scalia, J., dissenting).
39. *See* Newland, *supra* note 6, at 16.
40. *See, e.g.*, SLOTNICK & SEGAL, *supra* note 36, at 11.
41. *See id.* at 213.
42. *See* ZILIS, *supra* note 7, at 77-96.
43. *See id.* at 1-2.

direction more consistent with the work of Groseclose and Milyo,[44] there are opportunities to include those aforementioned reactions from politicians, think tanks, interested parties, and common citizens.

Of course, we do not suggest that all outlets treat their content in a similar way, or even that all authors of the same outlet do. At the same time, we also recognize that most prior bias measures are focusing on bias in all subjects, not just the Court. And while Ho and Quinn's measure is explicitly formulated based on editorials, not news content, we acknowledge that certain outlets do have reputations for ideological content.[45] In the era of "fake news," liberal bias has been broadly assigned to most outlets critical of the current administration, but the belief in widespread liberal bias is nothing new. On the other side of the aisle, there are certain informational sources that are almost universally considered to promote a conservative agenda—again, perhaps most directly through editorial or opinion based content. However, we again argue that the Court may be largely resistant or possibly immune, even among those outlets perceived to be ideologically driven, to such bias in coverage of its day-to-day activities.

This leads to our proposed test. We propose that news content, which we take to be the substance and words of the news articles themselves, should show an ideological bias if there is a connection between the news content and external readers' perceptions of their news slant. More specifically, we expect to see whether estimated measures of the ideological or partisan content correlate with known, external measures of media outlet ideology. This is essentially a similar test to the one proposed by Ho and Quinn,[46] but we test it on news content and not the editorial positions of newspapers. Importantly, we perform this test initially on Supreme Court stories of all cases and issue areas for one Supreme Court term. We then explore this relationship further by isolating conditions for when we think the media would be more or less likely to cover the Court in an ideological or partisan fashion. Specifically, we think likely conditions for ideological coverage would be: 5-4 decisions and in civil rights and liberties cases. We chose those conditions because they have been known to generate a significant amount of attention in the past and reside where much of the controversy surrounding the

---

44.   *See* Groseclose & Milyo, *supra* note 21, at 1191.
45.   *See* Ho & Quinn, *supra* note [12], at 355.
46.   *See id.*

Court seems to be. In contrast, most unanimous cases (9-0) do not generate widespread ideological or partisan concerns and should not encourage greater ideological or partisan coverage. We also test two other considerations, one stemming from a methodological consideration and the other from a journalistic consideration. Specifically, we examine the subset of stories that contain at least 200 words or more with the expectation that longer stories are more likely to contain ideological or partisan information. We also examine the subset of stories that have had Associate Press (AP) duplicate stories removed. We note a common journalistic practice for outlets is to use a basic AP story rather than their own content, so we expect more ideological or partisan content when the outlets put forth their own original content. Given these tests, if we fail to observe a significant relationship between the two constructs of interest, then it suggests to us support for the argument that there may be something unique about the Court and its coverage that buoys the idea that there are potential conditions when Court coverage does not have a clear ideological slant.

III. METHOD, MEASURES, AND DATA

To measure the ideological content in news stories covering Supreme Court decisions, we build our measure using the textual content of news stories. Specifically, we take a text-as-data approach[47] and scale stories from a range of news outlets. In this Article, as a first step, we propose using only a dictionary-based method, though we will use two different dictionaries—one theorized to comprise *partisan* words and the other comprised of *ideological* words.

Dictionary-based approaches are not conventionally conceptualized to be part of the scaling process. For example, Grimmer and Stewart do not even list dictionary methods as an option for scaling on their flowchart in providing an overview of text-as-data.[48] However, it should be noted that a dictionary approach is easily adaptable for this purpose.

Dictionary-based approaches have several benefits. First, the dictionaries are comprised of several words or phrases that the researcher knows, *ex ante*, to be theoretically related to the construct of interest. This is an advantage, at least initially, in that the

---

47. *See generally* Grimmer & Stewart, *supra* note 3.
48. *See id.* at 268.

researcher begins the process with a degree of validation for what the text-search process will return when it goes and counts the number of times words are used. In contrast, a completely unsupervised estimation procedure of scaling with text does not bring any *ex ante* information to the estimation process; thus, much effort has to be spent on the backend of the project validating whether the estimates are in fact what they were originally thought to be. The most common unsupervised method is the classic Wordfish approach, which is a text-scaling model based on relative word frequency that imposes no assumptions on the placement of words or documents in ideological space.[49] Nothing is inherently wrong with taking an unsupervised approach, but rather each approach can be seen as part of a tradeoff.[50]

Second, the dictionary-based approach highlighted here does not require us to bring any a priori assumptions about the *documents* themselves. Indeed, it is often the case that a supervised learning approach to scaling with text requires the researcher to assign a set of "reference" texts—predetermined places on the dimension that enable the "virgin" texts to be scored in relation to the reference texts.[51] This is the classic Wordscores approach.[52] In contrast, with a dictionary-based approach, the researcher gains the leverage of not having to place restrictive assumptions on certain *documents* because the researcher has instead made an assumption with respect to the words or phrases thought to represent the dimension of interest. The value of this tradeoff can be seen in that dictionary approaches: (1) do not lose the ability to make inferences about the small sample of reference documents in the analysis; and (2) the researcher is not forced to make an imprecise global judgment for the placement of a reference document, especially if one has not read all documents in the sample. In sum, a dictionary-based measurement approach has characteristics that make it similar to supervised learning approaches, but by making a different set of assumptions, it provides a benefit that a classic supervised approach cannot.

---

49.   Jonathan B. Slapin & Sven-Oliver Proksch, *A Scaling Model for Estimating Time-Series Party Positions from Texts*, 52 AM. J. POL. SCI. 705, 708 (2008).

50.   *See* Kevin M. Quinn et al., *How to Analyze Political Attention with Minimal Assumptions and Costs*, AM. J. POL. SCI., 209-10 (2010) (discussing this tradeoff in the context of topic models).

51.   *See* Michael Lavar et al., *Extracting Policy Positions from Political Texts Using Words as Data,* 97 AM. POL. SCI. REV. 313, 319 (2003).

52.   *See id.*

In contrast, of course, researchers are still imposing their own knowledge and biases in constructing the lists of words used in the dictionary. Thus, neither approach is necessarily better, and finding the appropriate method will ultimately depend on the goal(s) of the analysis. The goal of this Article is only to demonstrate the usefulness of dictionary-based approaches for measuring latent constructs in the legal world. The goal is not to argue that dictionary-based measurement strategies are superior.[53]

Third, while dictionaries are often criticized as blunt, they are extremely powerful given the relative ease with which they are implemented and applied. Furthermore, despite criticisms for their use outside the context for which they were developed,[54] dictionary-based approaches tend to be robust.[55] The concerns with dictionary-based approaches being used out of context have arguably led to critics being hypersensitive to the fear that dictionaries will make an overabundance of false negatives and false positives. In other words, the concern is that dictionary-based approaches will miss instances of phenomena that the researcher cares about (e.g., generate false negatives) and also inadvertently capture instances of word usage that are not consistent with the underlying meaning of the construct (e.g., false positives).

With the two dictionaries we chose below, we can assess the "out-of-context" criticism directly because we are using two dictionaries that were developed for a different context (e.g., governors' speeches and congressional speeches) and for a different time period (about ten years prior to the era we examine, the 2014 Term). If we find a nontrivial number of significant relationships between our dictionary measures of ideological and partisan content

---

53. To be sure, it is important to note how this dictionary-based approach is different from what Monroe, Colaresi, and Quinn discuss as feature selection, which examines a set of documents and selects words in a particular space. *See generally* Burt L. Monroe, Michael P. Colaresi & Kevin M. Quinn, *Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict*, 16 POL. ANALYSIS 372 (2008). What we are proposing here is trying to put documents in space by starting with some assumptions about a selection of words.

54. *See* Grimmer & Stewart, *supra* note 3, at 2.

55. For just one example, see the research on the dictionary-based program Linguistic Inquiry and Word Count [LIWC], which has been extended to many different domains and time periods. *See* James W. Pennebaker & Laura A. King, *Linguistic Styles: Language Use as an Individual Difference*, 77 J. PERSONALITY & SOC. PSYCHOL. 1296, 1297-98 (2000); Yla R. Tausczik & James W. Pennebaker. *The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods*, 29 J. LANGUAGE & SOC. PSYCHOL. 24 (2010).

and the facially valid external measures of media bias, then that should lend support to the claim of dictionary methods having some robustness. In fact, it suggests the possibility that if we were to develop word dictionaries specifically for the topic of focus, then our results may be stronger.

To provide as broad of coverage as possible, we use two different dictionaries. First, we use the ideological dictionary developed by Coffey in his examination of governors' ideology in State of the State addresses.[56] The second dictionary is a partisan word dictionary.[57] The list is comprised of the sixty most partisan two- and three-word phrases for both Democrats and Republicans from the 2005 Congress.[58] We chose both a partisanship and ideology dictionary to assess both constructs, even though in contemporary terms both of those will have a large degree of overlap. We discuss each of these in turn.

For our ideological dictionary, we use Coffey's list.[59] Coffey constructed lists of words for both liberal and conservative dimensions.[60] Specifically, he had lists of words for a redistributive dimension, a social dimension, and a government power dimension.[61] While there were also subscales for each of those three main dimensions, we focus only on those; and to simplify matters, we simply combine all of the dimensions and search for all liberal and all conservative words. A full list of these words is available in the appendix. The program then counts the number of times each dictionary word is present, and we are left with two variables— liberal words and conservative words. To construct an ideological dimension from those two variables, we use a simple ideological word polarization formula:

(liberal words – conservative words)/(liberal words + conservative words)

= ideological word polarization

---

56.   Daniel Coffey, *Measuring Gubernatorial Ideology: A Content Analysis of State of the State Speeches*, 5 St. Pol. & Pol'y Q. 88, 89 (2005).

57.   Matthew Gentzkow & Jesse M. Shapiro, *What Drives Media Slant? Evidence from U.S. Daily Newspapers* (Nat'l Bureau of Econ. Research, Working Paper No. 12707).

58.   *See id.*

59.   *See* Coffey, *supra* note 56, at 95.

60.   *See id.* at 91.

61.   *See id.* at 95.

For the partisan dictionary, we use the sixty most used partisan phrases for both Democrats and Republicans in 2005 Congressional speeches, as published in Gentkow and Shapiro's Table 1.[62] In general news stories, they found a relationship between outlet ideology and news content (recall, our focus is on stories about Court coverage).[63] The words (or phrases) used for these dictionaries are in Appendix A. Similarly, after the program counted the words in each dictionary, we are left with two variables—one for Democrat words and one for Republican words. We use a simple party word polarization formula similar to our ideological formula:

(Democrat words – Republican words)/(Democrat words + Republican words)

= party word polarization

We perform our text analysis and construct the dictionaries with R using the *quanteda* package.[64] Note: Both of the above polarization formulas standardize the word frequency by dividing by the total word counts of the total partisan and ideological words. We also try one other alternative way to standardize the word counts, and that is to transform each count by taking its log and using the above formulas.

For our external measure of ideology, we use a combination of two different measures. The first is the user ratings from mondotimes.com discussed earlier. They were previously used in both Ho and Quinn[65] and Gentzkow and Shapiro.[66] The second measure is from a second, similar user ratings site—AllSides.com.[67] Both measures use a 5-point scale, and they correlate quite well (r=.82). Hence, we "split-the-difference" between the two and take the average of both measures. It results in a scale that ranges from 1 to 5 (with increments every .5), where 1 is very liberal and 5 is very conservative. Some examples include Newsmax and Fox on the conservative end, the Daily Kos and the Daily Beast on the liberal

---

62. *See* Gentzkow & Shapiro, *supra* note 57, at 44-45.
63. *See id.* at 36.
64. *See* Kenneth Benoit & Kohei Watanabe, *An R Package for the Quantitative Analysis of Textual Data*, GITHUB, https://github.com/kbenoit/quanteda [https://perma.cc/B2RS-LZQ3] (last visited Feb. 19, 2018).
65. *See* Ho & Quinn, *supra* note 12, at 369.
66. *See* Gentzkow & Shapiro, *supra* note 57, at 47.
67. *See* ALLSIDES, https://www.allsides.com/unbiased-balanced-news [https://perma.cc/VZ4X-SKNY] (last visited Feb. 19, 2018).

end, and CNN and USA Today in the middle with a rating of 3. The ratings for all of the outlets we use are listed here:

1 = Beast, KOS, Huffington Post

1.5 = Salon, Boston Globe

2 = LA Times, Atlanta Journal Constitution, MSNBC, Washington Post, New York Times, ABC

2.5 = CBS, Reuters, AP, NPR, Chicago Sun Times, NBC

3 = CNN, Christian Science Monitor, Politico, USA

3.5 = Wall Street Journal, Chicago Tribune

4.5 = Washington Times

5 = Newsmax, FOX

For our measure of association, we rely on a simple yet straightforward measure of association, a pairwise correlation coefficient. For illustrative purposes, we also graph out each test, and some of these are shown below. Correlations are intuitive to use as a measure of association because of their ease of interpretation, with higher correlations bounded by 1 and -1, indicating a strong correlation, and correlations closer to zero representing a weak or nonexistent relationship. In terms of the direction of the correlation coefficient, given the formulas, where higher values represent more Democratic and liberal estimates, and given the external measures, where higher values represent more Republican and conservative estimates, we expect to see a negative correlation.

One weakness of relying on a correlation in this context is that it is sensitive to outliers when there are a small number of observations. In other words, because one or two observations can have disproportional influence, the correlation coefficient can be a conservative measure in terms of finding a significant relationship. Substantively, this means that there can be a significant relationship between the two constructs of interest, but that relationship may be drowned out by the measurement noise if one or two cases are outliers, giving a null reading.[68]

We utilize a dataset of more than 1,000 news articles on Supreme Court decisions in the 2014 term from twenty-six different

---

68. We should also note that correlations do not factor in a measure of uncertainty surrounding each outlet's point estimate. For example, an estimated location of an outlet may not fall on the imaginary diagonal that would create a perfect correlation between our measure and the external validated measure, but its confidence interval might.

news organizations. Important for our conceptualization of bias, we focused on news articles alone, not opinion or editorial content, and we limited our stories to appearing within three days of the announcement of the decision.

Admittedly, all of our stories, despite coming from sources that specialize in different media, come from online databases. We justify this decision based on the increasing trend toward online news consumption, and in the case of newspapers in our study, the content provided online and in print is most often identical. While we include most of the major outlets in American news (e.g., *The New York Times*, CNN, CBS News, etc.), our selected sources are geographically diverse as well, with sources from both coasts, the North, and the South respectively. Perhaps more importantly for the purposes of this Article, our sources span the perceived ideological spectrum, based on both the mondotimes.com scores and the second user-generated set of bias evaluations from AllSides.com.

To summarize our tests, we have two dictionaries (ideology and partisan), two weights for the words (frequency and log), and five different groupings for the stories (all stories, only stories with at least 200 words, removing AP duplicate stories, only stories covering 5-4 decisions, and only stories about civil rights and liberties decisions). We divided the stories into these different groupings because we expect them to influence the strength of the relationship between our estimate and the external indicator of outlet ideology. For example, we expect shorter stories to contain less partisan or ideological information simply because there is less information in shorter stories, and we expect outlets that use an AP wire story rather than produce their own content to contain less ideological or partisan information. Conversely, we expect that 5-4 decisions and civil rights and liberties cases will contain more ideological and partisan words. In total, we have twenty different possible correlations that will systematically test whether the news content covering Court decisions contains ideological or partisan content.

After the dictionaries are applied to the approximately 1075 texts, we then collapse the matrix by outlet, taking the mean of each outlet's score. This mean outlet score is then correlated with the external measure. This process is repeated for each grouping of texts.

IV. RESULTS

Table 1 contains the results from the twenty different correlation tests. Examining the column that contains correlations for all of the outlets (n=26), we see some interesting results. First, we see that the partisan dictionary achieved mild significance in almost all situations. To take one specific example, examine Model #2, which contains all stories and treats the words as frequencies: We see that the correlation is -.381 (p=.054). This suggests there is a mild degree of systematic partisan slant in the news stories covering Supreme Court decisions. To see what this looks like graphically, Figure 1 plots the correlation, with the x-axis being the external composite measure and the y-axis being the estimated dimension. It shows a mild negative relationship. Interestingly, some of the more liberal outlets (*Huffington Post* and *The Boston Globe*) and the more conservative outlets (*The Washington Times* and FOX) appear to be adding more noise to the correlation. Going further down Table 1 we see similar results for the section that uses only stories with at least 200 words or more, the section with 5-4 decisions, and the section that removes AP duplicate stories.

Table 1. Correlations of Dictionary-Based Estimates with External Measure of Ideology:

Model #	Documents	Word Weight	Dictionary	Correlation (n=26)	Correlation After Removing Influential Observations
1	All	Frequency	Ideology	-.082	-.467** (n=21)
2	All	Frequency	Partisan	-.381* (p=.054)	-.569** (n=24)
3	All	Log	Ideology	-.061	-.419* (n=21)
4	All	Log	Partisan	-.384* (p=.052)	-.572** (n=24)
5	docs with > 200 tokens	Frequency	Ideology	-.098	-.399* (n=22)
6	docs with > 200 tokens	Frequency	Partisan	-.355* (p=.075)	-.442** (n=25)
7	docs with > 200 tokens	Log	Ideology	-.072	-.376* (n=21)
8	docs with > 200 tokens	Log	Partisan	-.359* (p=.071)	-.446** (n=25)
9	Removes AP duplicates	Frequency	Ideology	.068	~

10	Removes AP duplicates	Frequency	Partisan	-.411** (p=.037)	-.531** (n=25)
11	Removes AP duplicates	Log	Ideology	.060	~
12	Removes AP duplicates	Log	Partisan	-.412** (p=.036)	-.537** (n=25)
13	Only 5-4 decisions	Frequency	Ideology	.257	~
14	Only 5-4 decisions	Frequency	Partisan	-.374* (p=.059)	-.530** (n=24)
15	Only 5-4 decisions	Log	Ideology	.194	~
16	Only 5-4 decisions	Log	Partisan	-.375* (p=.059)	-.529* (n=24)
17	Only civil rights & lib.	Frequency	Ideology	.449** (p=.021)	~
18	Only civil rights & lib.	Frequency	Partisan	-.129	-.456** (n=23)
19	Only civil rights & lib.	Log	Ideology	.511** (p=.008)	~
20	Only civil rights & lib.	Log	Partisan	-.144	-.572** (n=22)

**=p<.05, *=p<.1. The correlation column (second from right) represents the pairwise correlation with the public based ideology rating for all outlets (n=26). The far-right column shows the pairwise correlation coefficient after removing quasi-outlier observations (the column only shows a correlation if 5 or fewer observations could be removed to achieve a statistically significant correlation).
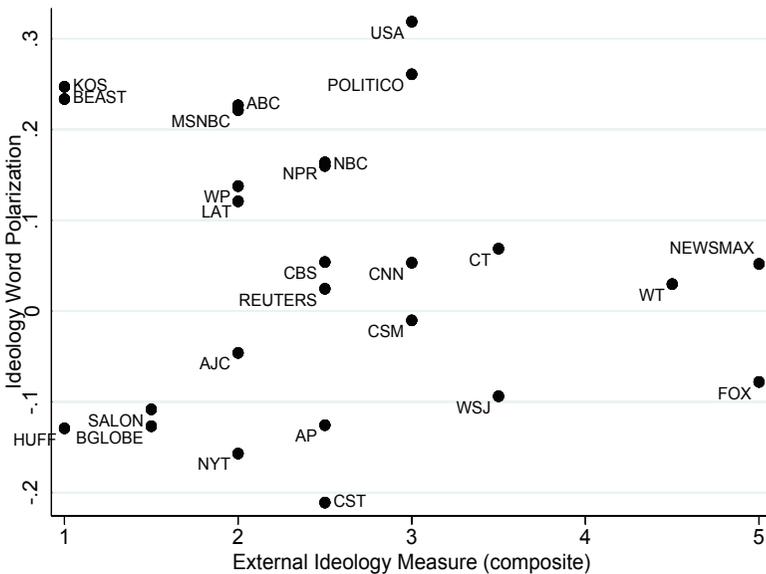
Figure 1. Partisan Word Polarization:
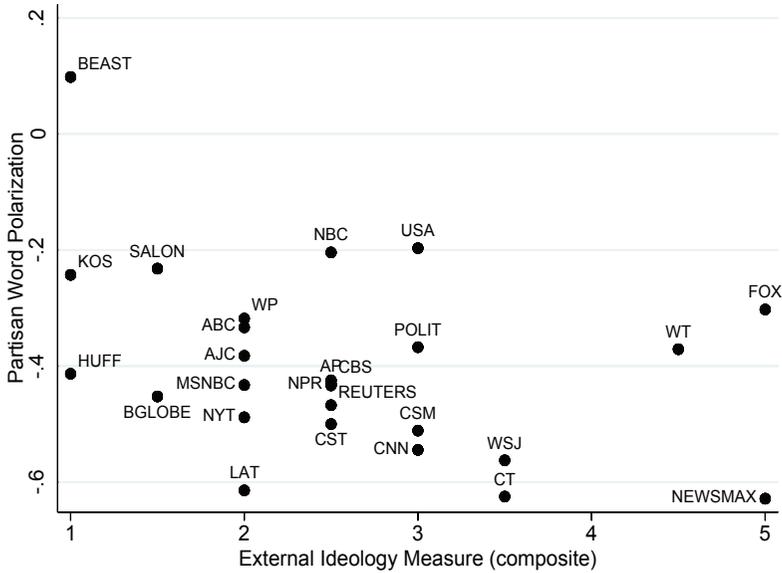
Note: Estimates from Model 2 of Table 1 (All documents, frequency, Partisan Dictionary). The Y-axis represents the estimated dimension based on the texts. Correlation is -.381 (n=26).

If we return to the top of Table 1, Model 1, perhaps the biggest surprise is the ideology dictionary coefficient's lack of magnitude and also lack of significance (r= -.082). We see this lack of significance in several other tests. In fact, the only place the ideology dictionary is significant is when it is in the wrong direction (Models #17 and #19), where we see that the ideology dictionary actually contains a positive sign, and in the case of civil rights and liberties cases. This suggests that under these limited circumstances, readers are potentially getting a slant that is opposite of what they might expect. However, we think something else might explain the lack of findings with the ideological dictionary.

We were worried about correlations being influenced too much by a small number of observations, so we examined some of the plots. Figure 2 shows the plot of Model 1 in Table 1. At first glance, there appears to be a potentially weak, negative correlation that is similar to the pattern found in Figure 1. However, if we look at the lower left corner of Figure 2, we see that there are some outlets that have a liberal reputation but were estimated to have more conservative content. Specifically, we see that *Huffington Post*, *The*

Boston Globe, *Salon*, *The New York Times*, and the *Chicago Sun Times* all were estimated to have more conservative content. Given this finding, we wanted to further probe the sensitivity of the initial results by seeing whether we could find a significant correlational relationship by removing as few observations as possible. We did this by examining a plot of each of the twenty-six outlets and removing up to five observations that were the most inconsistent with their reputations, leaving at least twenty-one, or 80%, of the outlets.

Figure 2. Ideological Word Polarization:

Note: Estimates from Model 1 of Table 1 (All documents, frequency, Ideology Dictionary). The Y-axis represents the estimated dimension based on the texts. Correlation is -.082 (n=26). If 5 outlets are removed (HUFF, SALON, BGLOBE, NYT, and CST), then the correlation changes to -.467 (see Table 1).

To be clear, this is not a formal test. Rather, it is a practical and quick way to assess the robustness of the findings. It is designed to see how sensitive each of the correlation coefficients are and if we might reach a different conclusion if we were to look at a major subset of the outlets. To place a limit on this data mining exercise, we only show the correlations for those models that we removed five outlets or fewer. The results for this are shown in the far-right column of Table 1. Now, when we look at approximately 80% of the

outlets (twenty-one outlets or more), we see that most tests give us a result that supports the finding that the news outlets contain some mild degree of slant across both partisan and ideological dictionaries. For example, we now see that Model 1 is significant, with the coefficient changing from -.082 to -.467 after removing just 5 observations. We see similar movements across all dictionaries.

Furthermore, to demonstrate how some relationships get even stronger, we only have to remove just a few observations. For example, Model 2 removes just two outlets and the correlation jumps from -.381 to -.569, which is a fairly strong relationship. These findings suggest that, while there are some outlets that appear to contain little or no ideological or partisan slant (or at least our estimate of their content is not consistent with their reputations), there are a large number of outlets that do contain partisan and ideological slant.

V. CONCLUSION AND IMPLICATIONS

The purpose of the Article was to demonstrate a relatively easy method for measuring a latent construct using a dictionary-based method that would be of interest to scholars of the empirical legal studies movement and legal analytics. We argued that the dictionary method was easy to implement; the researcher just has to come up with a list of words that represent the latent dimension. With it, our preliminary findings suggest that there is not always a clear connection between estimates of news content slant and external measures of media slant. To be sure, we obtained consistent results with the partisan dictionary but less so with the ideological dictionary. While we saw the partisan dictionary contained the most evidence of slant, the ideology dictionary seemed to be operating only on a larger subset of the outlets (80% of the outlets). These findings suggest that there is some credence to the claims that news content may be influenced by the happenings on the editorial board or that journalists are starting to use partisan content to construct their stories. This partly supports the arguments made in Ho and Quinn[69] and Gentzkow and Shapiro.[70] However, we would note that this relationship is far from clear and that, at least for one of our dictionaries, journalistic norms appear to play a role in constraining this relationship.

---

69.   *See* Ho & Quinn, *supra* note 12, at 372.
70.   *See* Gentzkow & Shapiro, *supra* note 57, at 64.

This has clear implications for the rule of law. Specifically, it suggests that individuals who must rely on information about the Court from media will not always receive information of a partisan or ideological flavor, despite prior suggestions. To be sure, in some instances they may receive a story that contains a mild slant. However, it is important to keep in mind that even if individuals only go to their preferred partisan or ideological source, the information they get will not always have a corresponding slant in relation to Court coverage. This provides some hope for advocates of democracy, and more directly the rule of law, in a time of great polarization.

In future iterations of this research, we see several avenues to explore further. We examined only one approach to estimating a latent dimension: a dictionary-based approach. To get a more complete answer to our substantive question, we could apply more classical approaches, such as an unsupervised learning approach with the Wordfish scaling program, or a supervised approach with the Wordscores scaling program. Going even further, one might incorporate Wordshoal,[71] which uses the unsupervised Wordfish framework but also accounts for the group-based nature of texts (by either author or case). Incorporating these alternative approaches should allow us to examine whether our conclusion is contingent on the method for estimating the ideological dimension.

Finally, we note that there is a distinction between content bias and coverage bias. In this Article, we are only trying to measure content bias, and we recognize the possibility that the more ideological bias may occur when an outlet chooses to cover some stories and not others. Perhaps one can model the selection mechanism to account for when an outlet chooses to cover a story that can be jointly modeled with a measure of news content. Regardless, we argue that examining the stories that did appear provides us with strong grounds for measuring the construct of partisan or ideological news content. This is something that can be used by both empirical legal scholars and those in the legal analytics field.

---

71. Benjamin E. Lauderdale & Alexander Herzog, *Measuring Political Positions from Legislative Speech*, 24 POL. ANALYSIS 374, 374-75 (2016).

APPENDIX A

Ideology of Outlet

To construct the ideology of the outlet, we relied on taking the average of two external measures: Mondotimes.com and allsides.com. Both sites use a 5-point scale and they correlate quite well—.82. As a simple way to "split-the-difference" between the two, we take the average of the two measures for our external measure of media ideology. This results in the below scale (note: we also have stories from Scotusblog.com and the Dallas Morning News, but we lose those observations because they both do not have a score to construct the average):

   1 = Beast, KOS, Huffington Post

   1.5 = Salon, Boston Globe

   2 = LA Times, Atlanta Journal Constitution, MSNBC, Washington Post, New York Times, ABC

   2.5 = CBS, Reuters, AP, NPR, Chicago Sun Times, NBC

   3 = CNN, Christian Science Monitor, Politico, USA

   3.5 = Wall Street Journal, Chicago Tribune

   4.5 = Washington Times

   5 = Newsmax, FOX

   Outlets not used because of no match with external valid indicator (Dallas Morning News, SCOTUSBlog.com) or too few stories (NY Post).

Coffey's (2005) Ideological Dictionaries (note: * indicates a wildcard and will search for all possible endings)

Conservative words (all dimensions combined): drain, irresponsible, out-of-control, out of control, rein-in, rein in, restrain*, runaway, spending, wasteful, balanced budget, cut spending, slash spending, unnecessary spending, cut, decrease, discourage, earn, incentives, lower, reduce, relief, slash, give back, tax, your money, attract*, contract*, roadblocks, friendly environment, enterprise zone, industry, compet*, entrepreneur*, private sector, right-to-work, right to work, individual*, responsibility, self-sufficient, private initiative, welfare rolls, able-bodied, able bodied, doctor-patient, frivolous lawsuits, liability insurance, government ownership, death panel, individual mandate, apprehend, assault, predator, castration, crime, deter, felon, firm, illegal, imprison, incarcerate, gangs, juvenile, lawless, lock, murder, parole, penal, prison, prosecute, punish, racket, rape, secure, sentence, steal, stolen, terror, three-strikes, three strikes, tougher, victim, weapon, convict, criminal, execute, offender, unlawful, violent, violence, bear arms, behind bars, crack down, death

penalty, gun rights, lethal injection, no parole, no tolerance, organized crime, law enforcement, almighty, atheism, character, Christian, corrupt, courtesy, disrupt*, duty, faith, heritage, honor, Lord, loyalty, moment-of-silence, moment of silence, moral, prayer, probity, tradition, religion, values, volunteer, wedlock, divorce, discipline, disciplinary, obscene, pornograph*, threat, parental notification, immoral, teen-pregnancy, illegal immigrant, quota, reverse discrimination, bloated, burden, bureaucracy, control, cripple, deregulation, excessive, hamper, inefficient, interfere, massive, regulate, red-tape, seize, stifle, regulatory reform, discour*, charter, account, flexibility, parochial, exam, performance, results, standards, strict, testing, vouch, merit pay, parental involvement, private schools, demand, video cameras, harvest, hunt, permits, fish, util*, private property, property rights;

Liberal words (all dimensions combined): earmarked, prudent, raid, underfunded, crucial, dangerous, necessary, rich, fair, wealth, equitable, workers, depend, fair growth, fair price, living wage, responsible growth, cycle, homeless, job-training, poor, poverty, health-care benefits, poor families, retirement benefits, single mom, single mother, working poor, access to healthcare, affordable, subsidize, CHIP, eligible, exploit, disable, guarantee, high-cost, insurance, necessity, enrollment, patient-rights, premiums, sue, uninsure, seniors, basic need, children's health insurance plan, insurance companies, welfare benefits, prescription drug benefit, elderly, medicaid, cruel, decriminalize, forgive, harass, harassment, hate-crimes, hate crimes, innocent, legalize, moratorium, non-violent, rehabilitate, child abuse, domestic abuse, domestic violence, drug counseling, overcrowded prisons, public defender, racial profiling, school violence, censorship, choice, choose, indoctrinate, privacy, repression, sex, unfair, union, civil unions, free speech, fundamental right, no place, personal choice, woman's right, African-American, blacks, discriminat*, freedom-of-religion, freedom of religion, hate, Hispanic, homosexual, human rights, injustice, inter-racial, legacy, lesbian, native-americans, prejudice, respect, stereotype*, unequal, dehuman*, diversity, women, equal, intrud*, liberal, minorit*, equal protection, religious rights, cigarettes, gun shows, gun violence, gun control, waiting period, danger, dysfunct*, exorbitant, freewheeling, gaming, gouging, irrespon*, manipulat*, overdevelop*, price-fix*, price-gouging, reasonable, reckless, require, poison, unrestricted, untested, basic necessity, public good, smart growth, universal health care, urban sprawl, bonus, noble, rebuild, reward, tuition, union, fair funding, classroom, public servants, school loan, teach, brownfield, carbon, co2, congestion, conservation, conserve, cycl*, desertification, dioxide, eco-system, energy-saving, energy saving, fragile, green, husbanded, monoxide, nature, opencast, over-use, ozone, planet, population, preserve, protect, re-use, rivers, smog, solar, species, stewardship, stream, toxic, tree, unleaded, warming, water, waterway, chemical, chimney, clean, deplet*, ecolog*, emission, environment, litter, pollut*, re-cycl*, wildl*, endangered species, future generations, acid rain, open spaces, hazardous.

Partisan Dictionary Taken from Gentzkow and Shapiro (2010) Table 1:

Democrat phrases: private accounts, Rosa Parks, workers rights, trade agreement, President budget, poor people, American people, Republican party, Republican Leader, tax breaks, change the rules, Arctic refuge, trade deficit, minimum wage, cut funding, oil companies, budget deficit, American workers, credit card, Republican Senators, living in poverty, nuclear option, privatization plan, Senate Republicans, war in Iraq, wildlife refuge, fuel efficiency, middle class, card companies, national wildlife, veterans health care, corporation for public broadcasting, cut health care, congressional black caucus, civil rights movement, VA health care, additional tax cuts, cuts to child support, billion in tax cuts, pay for the tax cuts, drilling in the Arctic National, credit card companies, tax cuts for people, victims of gun violence, security trust fund, oil and gas companies, solvency of social security, social security trust, prescription drug bill, Voting Rights Act, privatize social security, caliber sniper rifles, war in Iraq and Afghanistan, American free trade, increase in the minimum wage, civil rights protections, central American free, system of checks and balances, credit card debt, middle class families;

Republican phrases: stem cell, personal accounts, retirement accounts, natural gas, Saddam Hussein, government spending, death tax, pass the bill, national forest, illegal aliens, private property, minority leader, class action, border security, urge support, war on terror, president announces, cell lines, embryonic stem, human life, cord blood, tax relief, Chief Justice, action lawsuits, illegal immigration, human embryos, economic growth, date the time, increase taxes, food program, embryonic stem cell, circuit court of appeals, tongass national forest, hate crime legislation, death tax repeal, pluripotent stem cells, stem cells, housing and urban affairs, Supreme court of Texas, oil for food program, million jobs created, Justice Priscilla Owen, energy and natural resources, oil for food scandal, American Bar Association, global war on terror, private property rights, growth and job creation, hate crimes law, temporary worker program, natural gas natural, change hearts and minds, class action reform, Grand Ole Opry, global war on terrorism, Chief Justice Rehnquist, reform social security.