Recounting the Courts? Toward A Text-Centered Computational Approach to Understanding the Dynamics of the Judicial System

Michael Evans
Department of Government & Politics
University of Maryland
mevans@gvpt.umd.edu

Wayne McIntosh
Department of Government & Politics
University of Maryland
wmcintosh@gvpt.umd.edu

Cynthia L. Cates
Department of Political Science
Towson University
ccates@towson.edu

Jimmy Lin
College of Information Studies
University of Maryland
jimmylin@umd.edu

Paper prepared for delivery at the 2005 annual meeting of the Midwest Political Science Association, Chicago, Illinois, April 7-10, 2005

Recounting the Courts? Toward A Text-Centered Computational Approach to Understanding the Dynamics of the Judicial System

Abstract

This paper explores the potential uses of computational linguistics techniques for analyzing Supreme Court briefs and opinions. To do so, we focused on advocacy documents associated with the two recent University of Michigan affirmative action cases (*Gratz v. Bollinger* and *Grutter v. Bollinger*). The cases attracted more than one hundred *amicus* briefs, which provide a rich textual database for such an exploratory study. The goal of our preliminary work is to model the linguistic contents of the arguments presented by the petitioners and respondents, as captured in the original litigants' briefs and the *amici* briefs submitted in these two cases. In particular, we are interested in the types of words and phrases used by both sides in order to forward their arguments. These linguistic cues may provide us with insight into the policy and ideological inclinations of the parties involved. We utilize and compare two analytical methods, an adaptation of the *Wordscoring* technique first developed by Laver, Benoit, and Garry (2003) and a Naïve Bayes' classifier to identify commonalities in documents and group them accordingly. We find the methods to be quite competent at detecting *amici* brief positions, clustering petitioner and respondent briefs into well-spaced separate normal distributions. Additionally, we find it quite useful as an aid to qualitative content analysis. We identify distinctive rhetorical styles utilized by the respondents and petitioners and suggest how this type of analysis can improve our understanding of how and why different groups, ideologies, actors, interests, and the like, conceptualize issues.

1. Introduction[1]

The legal system is an elaborate and intricate means of communicating, processing, and transferring information. It consists of agents (e.g., lawyers, judges, litigants, interest groups) whose behavior is affected by a range of influences (e.g., political ideologies, historical precedent, current political and economic context, power differentials). The law, as expressed in judicial opinion, is never static, but constantly evolves to address emerging challenges and opportunities. It leaves textual records that explain to litigants who won and why. In turn, these records are referenced by future litigants and judges. As such, the judicial process is founded on rhetorical power, expressed through text, and organized in a hierarchical institutional structure where past decisions inform present conflicts. Because critical aspects of the judicial process are grounded in documents produced by judges, counsel, and third parties, analysis of these texts may help scholars better understand the workings of this complex social system.

[1] We would like to acknowledge a number of colleagues who have contributed in one way or another to this project. Ken Cousins, Ken Benoit, Mark Kessler, Irwin Morris, and Geoff Layman commented on earlier drafts; while Rebecca Thorpe, Marlaine White, and Amy Hershkovitz provided important research support.

This work presents a preliminary attempt at applying computational techniques to automatically analyze legal documents within the framework of text classification.

In computational terms, we can think of "judicial output" as a sequential, time-ordered text stream, contributed by multiple authors. Inasmuch as the content of the documents within this stream directly reflect the dynamics of the system and relationships among participants, advances in our ability to automatically store, search, classify, and retrieve textual information offer enormous opportunities to improve our understanding of judicial processes. In recent decades, information retrieval researchers have developed technology to manipulate both static and dynamic document collections. In addition, computational linguists have developed tools to automatically analyze textual documents at the lexical, syntactic, and semantic levels. We believe that now is an opportune time to apply some of these computational tools to study interesting problems in political science. Though judicial scholars have long noted the importance of opinions and briefs in developing an explanation for the behavior of key actors (i.e., justices, litigants and *amici curiae*), practical obstacles have limited efforts to analyze the content of large samples of legal text. Advances in text processing by computers makes large-scale analysis of legal documents not only a possibility, but an exciting opportunity.

This paper is organized as follows: in Section 2, we describe our computational approach to the analysis of legal texts using the framework of text classification, a well-studied problem in computational linguistics and information retrieval. Section 3 offers additional background on the judicial system relevant to this research. Previous efforts to apply quantitative approaches to the analysis of legal texts are discussed in Section 4, followed by a report of our own preliminary work in this area, focusing on the *amici* briefs submitted in the 2003 University of Michigan affirmative action cases. The paper concludes with a brief discussion of future research and impact.

2. A Computational Model of the Judicial Process

In a political system marked by tension between democratic rule and the rule of law, the question of how the US Supreme Court contributes to the creation, destruction, and perpetuation of binding legal doctrine is of clear importance. One simplistic model of the American legal system suggests a minimal role for the Court as an agent of legal change. Elected officials in the legislative and (to a lesser extent) executive branches are said to be the sole sources of legal rules and regulations within a context of constitutionally defined rights, duties, and countervailing powers. The Court, on the other hand, is understood to be an agent of stability, ensuring that democratic demands are tempered by constitutional limits—that is, by the rule of law. Most careful observers of the judicial process do not find this model convincing, but rather, argue that justices seek to influence policy, and in the process, effectively contribute both to change and continuity in the law. Although our understanding of judicial behavior has improved considerably over the past century, there is still much debate about the degree of influence the Court has over legal developments. To what extent do justices deviate from the norm of *stare decisis*? In what way do they do so? Do external factors limit their ability to effect legal change? Are justices persuaded by litigants or third parties through the arguments presented in their briefs? Or do they behave as relative "free-agents," voting and writing "according to their ideological attitudes viz. a viz. the facts of [each] case" (Segal and Spaeth 2002)? Answers to these questions can improve our understanding of the dynamics of this complex human system.

Traditionally, legal research has faced a trade-off between large scale quantitative inquiries focused on "thin" observations (e.g., voting records, participation, coalition size, length of legal documents), and smaller-scale, close readings of legal texts. This inverse relationship between breadth and depth has limited our ability to observe, measure, and analyze change in legal texts over time. With techniques drawn from computational linguistics and information retrieval, it becomes possible to analyze large quantities of judicial texts—essentially rendering large-scale content analysis feasible.

Although the immediate concern of parties to a legal conflict is, of course, who wins, judicial outcomes involve much more. As Shapiro (1968, 39) asserts "[T]he opinions themselves, not who won or lost, are the crucial form of political behavior by the appellate courts, since it is the opinions which provide the constraining directions to the public and private decision makers who determine the 99 percent of conduct that never reaches the courts." While the Supreme Court may determine which side prevails today, the language articulated in its opinions lives on, influencing those charged with implementation, subsequent judges and justices faced with interpretation, and advocates who craft responses. Accordingly, our fundamental assumption is that understanding these documents is the key to understanding the dynamics of the legal system as a whole.

We conceive the judicial process as a time-ordered sequence of documents of particular types (e.g., judicial opinion, *amicus curiae* brief, and so on) about a particular topic (e.g., free speech, federalism, and so on). These documents tend to cluster temporally and topically in that multiple texts on the same topic appear in very close temporal proximity (all the documents regarding a particular case). Multiple clusters are generally separated by longer time spans, reflecting the evolution and development of an issue through legal history. We treat authors as rational agents, motivated by ideological and policy preferences, expressed as observable behavior (e.g., votes, written opinions, and briefs). Although these behaviors may vary by situation, we assume that the underlying motivations remain constant; that is, we assume that authors are consistent across issues and time with respect to their internal states (i.e., political ideology or worldview) that cannot be directly observed. The content of each document is directly determined by its type, the author's preferences, and by previous texts on the same topic.[2]

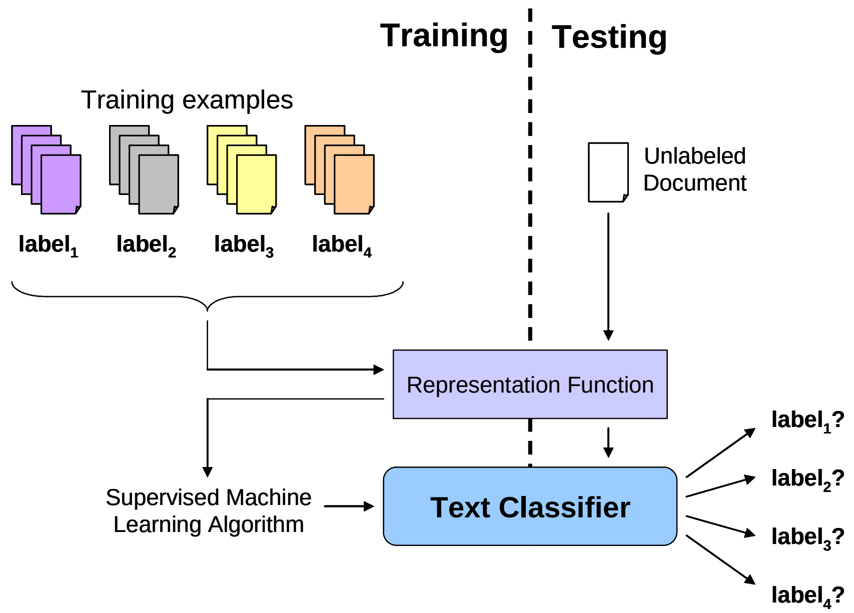
Our computational approach to the analysis of legal documents is cast within the framework of text classification using machine learning techniques. We first briefly describe the text classification problem, and then outline how it can be employed to explore interesting questions in political science.

Text classification is a well-studied problem that lies at the intersection of computational linguistics and information retrieval (Lewis 1992; Brill and Mooney, 1997; Knight 1999). This problem can be intuitively described as the task of automatically sorting "items" into "bins." In our domain, these "bins" (called labels, categories, or classes) could correspond to any directly or indirectly observable characteristic, such as political ideology, issue bias, or voting behavior. Our goal is to

[2] Other causal factors, such as (internal and external) strategic considerations, case facts, and (formal and informal) institutional change, can also influence the opinion writing process.

label previously unseen documents correctly, where the labels are drawn from a finite number of alternatives. We adopt a machine learning approach to this problem (Mitchell 1996; Sebastiani 2002), where different algorithms are applied to automatically "learn" characteristics that distinguish one type of text from another based on examples that have been labeled *a priori*.

Figure 1: The Machine Learning Approach to Text Classification.



As Figure 1 schematically shows, the machine learning approach to text classification can be divided into two phases.

In the **training** phase, the system is presented with correctly labeled documents from which to learn (e.g., judicial opinions annotated by the political ideology of their authors). Typically, these labels are manually assigned by humans who have already analyzed the text according to a theoretically-grounded classification based on the research question being explored. In many cases, this analysis has already been performed, and is often stored as metadata attached to the documents. Since computers cannot "understand" documents in the same way humans do, "learning" takes place at the level of abstract models automatically generated by the representation function, usually in terms of "features" extracted from the training examples. A feature can be any quantifiable characteristic of the text, for example, the presence of certain words. These features, which can be thought of as a "digest" of the text, and the pre-assigned labels together serve as the input for the machine learning algorithms that will be used to train the text classifier.

In the **testing** phase, the trained classifier is presented with new unlabeled documents (naturally, previously unseen in the training examples), and the computer's task is to correctly assign labels consistent with the training examples. The system's performance is based upon the accuracy of the classifier (i.e., the proportion of the computer-assigned labels that were, in fact, correct). Accuracy measures can be further broken down in terms of a two-by-two contingency table: true positives, true negatives, false positives, and false negatives.

How does the text classification task relate to questions in political science? We illustrate with a few examples: Let us consider the interesting problem of predicting justices' voting behavior. We assume that the decision is, at least in part, based on the contents of briefs filed by the various involved parties. Vote prediction can be formulated as the task of predicting a justice's reactions to the arguments presented in a particular brief (i.e., agreement or disagreement): this translates into a standard binary (two-class) classification problem. In this particular example, the training examples would be previous briefs read by the justice and the outcome in those cases (which we already know). As an alternative, we could also label texts based on ideology and attempt to predict voting patterns by associating those labels with known ideological biases of the justices (e.g., a judge known to be conservative would be more sympathetic to a brief that advocates a conservative position). The ideological inclinations of justices could in turn be derived from the opinions authored by the justices themselves. Ultimately, multiple sources of evidence could be combined to accurately model the behavior of any agent within the legal system.

As another example, suppose we wish to explore the evolution of a particular idea, say, liberal views on labor issues, as expressed in court opinions. A necessary prerequisite to such an analysis would be first to identify those documents that both reflect liberal thinking and discuss labor issues—again, this is a text classification where a text classifier can be employed to automatically label large amounts of data (e.g., with their ideological orientation and type of issue discussed).

In this study, we describe our attempt at applying text classification to the task of detecting the policy position of *amicus* briefs. This specific problem is situated within an exploration of quantitative techniques for modeling legal arguments. As we will show, our results validate a text-centered computational approach to the study of the legal system.

3. Background

Social scientists who study the American judiciary generally concede that judges attempt to promote their policy preferences through the cases presented to them for decision (e.g., Segal and Spaeth 2002). In contrast to the simplistic model described earlier, most recent judicial analyses generally assume that case outcomes reflect the underlying ideological or policy inclinations of judges. However, a half-century ago the prevailing thought was that judges merely apply objectively derived principles of law as established by precedent, logically extending them to contemporary controversies. According to this "legalist model" of judicial behavior, judges are guided by the doctrine of *stare decisis*, which strictly binds them to apply the law embedded in previously decided cases (e.g., Levi 1949). Challenging conventional wisdom, Pritchett (1941) noted that dissents at the U.S. Supreme Court level, historically minimized by institutional restraints, had risen by 1940 to nearly half of all cases receiving full treatment. Understanding such decisional divisions, he argued, requires us to consider factors such as ideological leanings and value preferences. Prichett's (1948, 1954) analysis of

Supreme Court behavior, revealing voting blocs across a range of issue areas, has influenced several generations of judicial scholars since (e.g., Schubert 1959, 1965, 1974; Rohde and Spaeth 1976; Segal and Spaeth 1993, 2002). However, though justices do appear to express fairly consistent points of view (as indicated by their voting records), their range of choices is somewhat limited by precedent[3] (Brenner and Stier 1996; Brisbin 1996; Knight and Epstein 1996; Segal and Spaeth 1996; Segal and Spaeth 1996; Songer and Lindquist 1996; Spaeth and Segal 1999; Spaeth and Segal 2001) and other institutional constraints (e.g., Richards and Kritzer 2002). Moreover, they are not immune to mutual influence or to arguments of the advocates who populate their docket (e.g., Murphy 1964; Howard 1968; Rohde 1972; Epstein and Knight 1998; Wahlbeck, Spriggs et al. 1998; Spriggs, Maltzman et al. 1999; Epstein and Knight 2000; Maltzman, Spriggs et al. 2000; Caplan 1987; Maltzman and Wahlbeck 1996; Johnson 2001; Johnson 2003; Johnson 2004). Furthermore, some suggest their decisions are influenced by the preferences of key Congressional and Presidential actors (Gely and Spiller 1990; Eskridge Jr. 1991; Spiller and Tiller 1996; Segal 1997; Bergara, Richman et al. 2003; Johnson 2003).

It is well known that inferring ideological and policy preferences from votes presents tricky analytical problems. Although justices do appear to vote consistently on particular issues, in the absence of external supporting evidence, such models can be critiqued as tautologies. Acknowledging that "one cannot demonstrate that attitudes affect votes when the attitudes are operationalized from those same votes," Segal and Cover (1989, 558) offer an alternative *a priori* measure, using pre-confirmation newspaper editorials characterizing individual justices' ideological perspectives as a proxy.[4] The Segal-Cover index has proven to be a fairly useful predictor of votes in civil rights and civil liberties cases (e.g., Segal and Cover 1989; Segal et al. 1995) but less satisfactory beyond those issues (Epstein and Mershon 1996).[5]

Moreover, the spreading occurrence of splintered opinion outcomes, with one or more concurrences, separate dissents, partial concurrences, and dissent further devalues the usefulness of vote tally as a meaningful indicator to differentiate policy preferences among justices. The norm of consensus began to break down during the Hughes Court of the 1930s, and the growing expectation is to see a multiplicity of opinions in response to any given case (Caldeira and Zorn 1998; O'Brien 1999; Epstein et al. 2001; Post 2001). Indeed, O'Brien observes that during the Burger Court era, in both *Furman v. Georgia* (408 US 238, 1972) (death penalty) and *U.S. v. New York Times* (403 US 713, 1971) (Pentagon Papers), "the justices produced ten opinions—a *per curiam* opinion, announcing the Court's decision, six concurrences, and three dissents" (1999, 108).[6]

[3] Sunstein (1999, 42) posits that prior case law culls "certain arguments from the legal repertoire [which] simplifies analysis. Most of the important constraints on judicial discretion [in interpreting the Constitution] come not from constitutional text or history, but from the process of grappling with previous decisions."
[4] Cf. Danelski (1966), who uses past speeches of justices as an indicator; Tate and Handberg (1991); Ulmer (1986), who attempts to correlate background characteristics with policy and ideological preferences.
[5] Epstein's and Mershon's analyses also suggest that the index may have better predictive value during the Burger Court than for earlier or later Courts.
[6] In 1930, only about 8% of the Court's cases saw a dissenting opinion, which was representative of that era. By 1953, the comparable figure was 56%, which is a fairly typical dissent rate for the last half-century.

Although the rising cacophony of judicial opinions can create daunting problems for those charged with making sense of them, we view this as an exceptional research opportunity.

At the same time, third party participation in Supreme Court litigation has been on the rise since the early 1950s. Some cases draw more *amici curiae* briefs than others, but more than 90% of all cases heard by the Court now attract at least one (Epstein and Knight 1999). We can assume that *amici* hope to influence the Court's decision and thus advocate a specific policy preference (e.g., Caldeira and Wright 1990; Epstein and Knight 1999; Hansford 2004; Spriggs and Wahlbeck 1997). Although we have reason to believe that their presence can be influential (e.g., Caldeira and Wright 1990; Segal 1991), uncovering evidence that their *arguments* have an impact on the law has proven quite challenging (see e.g., Songer and Sheehan 1993; Spriggs and Wahlbeck 1997; Epstein and Knight 1999). Statistical analyses in this strand of research typically focus on discrete data points and small sample sizes (e.g., unearthing patterns in registered appearances before the court and final decisions, and inferring influence on legal developments from observed correlations).

Characterizing the opinions expressed in legal documents—and inferring authors' underlying motivations—is a time-honored tradition among legal scholars, albeit one with clear limitations. Legal texts tend to be lengthy and dense, presenting serious challenges to even the best doctrinal specialists and historians. Furthermore, content analysis by humans is inevitably subjective. Even individual scholars find it difficult to maintain consistency when coding complex documents, particularly when the objective is to compare multiple documents. This problem is even graver for team-based approaches (e.g., Carmines and Zeller 1979).[7] In addition, the labor-intensive nature of hand-coding induces serious limits to the scale at which such approaches can even be attempted. In recent decades, information retrieval researchers have developed techniques for efficiently storing and retrieving texts at a large scale (Frakes and Baeza-Yates 1992; Salton 1989), and computational linguists have developed increasingly sophisticated statistical algorithms to analyze content (Manning and Schütze 1999). Nevertheless, computer systems may still take advantage of specific human-engineered domain knowledge (e.g., pre-coded dictionaries consisting of words and phrases theoretically linked to analytical dimensions of the coder's choosing). This allows researchers to have the best of both worlds: detailed analyses of legal documents at a large scale.

4. Previous Work

Quantitative techniques have been previously used by scholars in other disciplines to analyze text. In their seminal work, Mosteller and Wallace (1964) identified the author of twelve of the *Federalist* papers, whose authorship had been considered ambiguous, by applying Bayesian inferential statistics and treating words as data sequences. They first

---

In addition, the percentage of cases with at least one separate concurring opinion has steadily increased. In 1930 it was only about 1%; by 1953 it had reached 16.7%, and in 2001 the concurrence rate was 37.7% (Epstein et al. 2003, pp.209-220).

[7] Also see the reliability analysis James Gibson conducted as part of the Supreme Court Database Project (Gibson 1997).

developed two “function word” lists based upon patterns of usage in texts known to have been written by Hamilton and Madison. These were then used to assess the authorship of the contested documents, now generally accepted as being written by Madison (see, for example, Martindale and McKenzie 1995, who utilized content analysis and function word techniques). Mosteller and Wallace essentially treated the authorship problem as a binary classification task, using texts with known authorship for training. Other scholars have applied computational techniques to assess the authorship of such texts as books of the New Testament (Morton 1993; Morton and McLeman 1980; Kenny 1986), the Book of Mormon (Larsen et al. 1980), and Shakespearean era plays (Foster 1989). Among the best known examples of what has come to be known as “forensic linguistics” was Donald Foster’s attempt to identify the author of *Primary Colors*, a thinly veiled political novel based on the 1992 Clinton presidential campaign. Published “anonymously” in 1996, the book immediately set off endless speculation about who had written it. Foster collected a database of documents written by likely candidates, and after analyzing them using a computer and comparing the results to an electronic copy of the book, he correctly identified Joe Klein (Foster 2000). After weeks of repeated denials, Klein finally admitted his authorship.

More recently, social scientists Laver, Benoit, and Garry (2003) experimented with several quantitative techniques before developing a “Wordscoring” method to identify ideological differences expressed in the platforms of the three major political parties in the UK. Beginning with documents whose positions (e.g., liberal versus conservative) are known, the researchers juxtaposed new documents with the originals, comparing the frequency at which the authors used particular words to express their positions. “Wordscores” can be calculated based on the probability that individual words are found in the reference (training) texts. Based on this, “textscores” can be computed to measure the similarity of unread “virgin” (test) texts to the reference documents.

Despite these previous works, quantitative analytical approaches are generally unusual in the social sciences and humanities. We believe that computational techniques provide an entirely new approach to tackling traditional problems. Legal briefs and judicial opinions provide an excellent environment for exploration and study—techniques from information retrieval and computational linguistics may provide the keys that will allow us to better understand the complexities of the legal system.

5. Analyses

In this section we assess the arguments associated with two cases heard and decided together during the 2002 term of the U.S. Supreme Court, *Gratz v. Bollinger*, 539 U.S. 244 (2003), challenging the University of Michigan undergraduate affirmative action admissions policy, and *Grutter v. Bollinger*, 539 U.S. 982 (2003), challenging the University of Michigan Law School admissions policy. The two cases attracted 104 *amicus curiae* briefs and yielded 13 opinions from the Court, with one majority (6-3) finding in favor of Gratz and thus invalidating the undergraduate admissions system, and another majority (5-4) finding against Grutter to uphold the law school admissions process. While the numbers here are quite large, and there are plenty of votes to count, the multiplicity of briefs offer a potentially rich textual database that we exploit with the use of analytical techniques associated with computational linguistics.

The goal of our preliminary work is to model the linguistic contents of the arguments presented by the petitioners and respondents, as captured in the original litigants' briefs and the *amici* briefs submitted in these two cases. In particular, we are interested in the types of words and phrases used by both sides in order to forward their arguments. These linguistic cues may provide us with insight into the policy and ideological inclinations of the parties involved. It seems fairly obvious that petitioners are more likely to draw upon different words than respondents, but determining the exact distribution of salient words is beyond the reach of manual content analysis. This is an area where automated techniques for analyzing textual content can be fruitfully applied. This study specifically compares the "Wordscoring" method proposed by Laver, Benoit, and Garry (2003) with a Naïve Bayes' classifier within the text classification framework we described earlier.

Specifically, we examine the problem of classifying *amici* briefs as either supporting the petitioners or the respondents—this can be viewed as an attempt to automatically detect the policy position of the briefs based on their linguistic content alone (without knowledge of their declared positions). This task was chosen for a variety of reasons. From a technical perspective, because "ground truth" is known, evaluation of our classifier is very straightforward: we simply compare the computer assigned labels to the known true labels. From an intellectual perspective, good performance in this task validates our general methodology for casting political science questions in the framework of text classification. The same approach can certainly be extended to predict behavior which we cannot directly observe or events that have yet to happen. We hypothesize that the substance of the arguments presented in the litigants' and *amicus* briefs can be captured in terms of their linguistic content using computational models. Our preliminary results appear to support this hypothesis.

5a. Case Background. The Supreme Court's 2002-03 Term featured a number of high profile cases. Much anticipated were rulings on gay rights (*Lawrence v. Texas*, 539 U.S. 558), three strikes laws (*Lockyear v. Andrade*, 538 U.S. 63 and *Ewing v. California*, 538 U.S. 11), internet indecency (*U.S. v. American Library Association*, 539 U.S. 194), cross burning (*Virginia v. Black*, 538 U.S. 343), racial gerrymandering (*Georgia v. Ashcroft*, 39 U.S. 461), and states' rights (*Nevada Department of Human Resources v. Hibbs*, 538 U.S. 721). But, the most awaited decisions of the Term – those widely believed to have the greatest potential impact – were two affirmative action cases, *Gratz* and *Grutter*.

Gratz and *Grutter* were challenges to the use of race in admissions at the University of Michigan and the University of Michigan Law School, respectively, during the 1990s. In one sense, then, the cases were both parochial and current. In more important respects, however, the disputes were neither, reaching back some 25 years and across the whole landscape of higher education.

The progenitor, of course, was *Regents of the University of California v. Bakke* (438 U.S. 265, 1978), decided a quarter century before by a divided Court. Indeed, *Bakke* generated three major opinions. Justice Brennan, joined by Justices White, Marshall, and Blackmun, concluded that neither Title VI of the Civil Rights Act nor the 14th Amendment "bar the preferential treatment of racial minorities as a means of remedying past societal discrimination to the extent that such action is consistent with the . . .

amendment" (Id., at 328, Brennan, J., concurring in the judgment in part and dissenting in part). On the other side, Justice Stevens, along with Chief Justice Burger and Justices Stewart and Rehnquist, argued that the case could be settled on the basis of Title VI alone which, they contended, clearly prohibits racial preference in any program receiving federal funds (Id at 408-421, Stevens, J., concurring in the judgment in part and dissenting in part).

Thus, Justice Powell, straddling the middle, became the voice of the Court, finding that race may be used as one of a number of factors in university admissions programs, but that the UC Davis policy amounted to an unconstitutional racial quota (Id at 297ff).

The decision forced universities throughout the country to rethink affirmative action programs, many attempting to model the Harvard program cited approvingly by Justice Powell (Id at 316). The Court's fractured mandate, however, meant, inevitably, that more challenges would follow. And, indeed, they did, picking up steam particularly in the 1990s.[8] The fact that different Circuits were coming to very different conclusions, ultimately forced the High Court's hand. To settle the issue, it chose the University of Michigan, which employed one method of achieving diversity in its undergraduate admissions and another in its law school admissions.

5a(1). *Gratz v. Bollinger*. The University's undergraduate program relied on a point system. Applicants could be awarded a possible 150 points, with plus factors including "high school grades, standardized test scores, high school quality, curriculum strength, geography, alumni relationships, and leadership" (539 U.S. 244, at 253). An additional positive was race; in particular African Americans, Hispanics, and Native Americans, were considered "under represented minorities" by the University. Members of these groups received 20 points toward admissions. Gratz, a Caucasian student and Michigan resident denied admission in 1997, filed a class action against the University. The District Court ruled for the defendant, finding "the educational benefits flowing from a racially and ethnically diverse student body [to be] a sufficiently compelling interest to survive strict scrutiny" (*Gratz v. Bollinger*, 122 F. Supp. 2d 811, at 824 (E.D. MI, 2000)). Moreover, it judged the school's ongoing admissions program to be narrowly tailored, thus meeting "the requirements set forth by Justice Powell in Bakke" (Id, at 831). While the *Gratz* appeal to the Sixth Circuit was pending, the Supreme Court opted to go ahead and hear the case with another that the Circuit had recently decided in favor of the University of Michigan Law School.

5a(2). *Grutter v. Bollinger*. For its part, the law school, notably, one of the nation's most elite, employed a less formulaic admissions policy. As with almost every other law school, major determinants were LSAT scores and undergraduate GPA. In

[8] See e.g., *Johnson v. Board of Regents of the University of Georgia*, 263 F.3d 1234, U.S. Court of Appeals for the Eleventh Circuit (2001); *Hopwood v. Texas*, 236 F.3d 256, U. S. Court of Appeals for the Fifth Circuit (2000); and *Podberesky v. Kirwan*, 38 F.3d 147, U. S. Court of Appeals for the Fourth Circuit (1994). See also, California Const., art. I, § 31 (Proposition 209) which provides, "The state shall not discriminate against, or grant preferential treatment to, any individual or group on the basis of race, sex, color, ethnicity, or national origin in the operation of public employment, public education, or public contracting."

addition to these numeric variables, the admissions committee was allowed to entertain certain "soft" variables, "like the enthusiasm of the recommenders, the quality of the undergraduate institution, the quality of the applicant's essay, residency, leadership and work experience, unique talents or interests, and the areas and difficulty of undergraduate course selection" (*Grutter v. Bollinger*, 288 F.3d 732, at 736 (6th Cir., 2002)). Taking these factors into consideration, the law school would sometimes admit students with relatively low scores if it had "good reason to be skeptical of an index score based prediction . . . [or, if the student in question] may help achieve that diversity which has the potential to enrich everyone's education and thus make a law school class stronger than the sum of its parts" (Id). According to the law school, diversity factors might include "an Olympic gold medal, a Ph.D. in physics, the attainment of age 50 in a class that otherwise lacked anyone over 30, or the experience of having been a Vietnamese boat person" (Id). Although attaching no number to it, the school admitted its desire to achieve and maintain a "critical mass" of underrepresented minorities, "a number sufficient to enable under-represented minority students to contribute to classroom dialogue without feeling isolated" (Id, at 737). The appeals court ruled in favor of the law school plan.

An indication of the perceived importance of any case is the number of *amici* inspired to join in the fray. On that basis alone, *Gratz* and *Grutter* could be considered unusually consequential. A total of 104 *amici* briefs were filed with the Supreme Court, 19 in support of the petitioners, 79 backing the respondents, and 6 claiming to be neutral. Thus, if *amici* provide the Court with information, the Justices had data to spare.

5b. *Wordscores* Analysis. As a baseline, we employed the methodology developed by Laver, Benoit, and Garry (2003), using their *Wordscores* program.[9] The process begins with selection of "reference" (training) documents, written with an understood point of view along the dimension of interest (e.g., ideology, policy issue field, and the like). The *Wordscores* program then generates a word frequency matrix for every word in the reference documents.[10] Based on the relative frequencies of each word in the reference documents and the values assigned to those documents, wordscores are then calculated to quantify the association between words and either document. For example, let us assume reference document D_1 is assigned a value of -10 and reference document D_2 is assigned a value of 10. Let us further suppose that word w_{20} is used 8 times out of 3000 words in D_1, and 150 times out of 5000 words in D_2. Since w_{20} is used far more frequently in D_2 than in D_1, it will receive a score closer to 10 than to -10, suggesting that the word is more indicative of the position of D_2 along the given dimension than that of D_1. Finally, textscores are computed to compare proximity to the reference documents among one or more unread, uncharacterized "virgin" text(s); these texts comprise our test examples. The score given to each virgin text is simply the average of all wordscores for all scored words within the text. In our example above, all things held equal, a virgin text that includes w_{20} at a high frequency would receive a

[9] The *Wordscores* program is written to work within *STATA* statistical software (version 7 or 8). The authors have established a *Wordscores* Main Page (http://www.politics.tcd.ie/wordscores/) that includes a free download of the *Wordscores* program, miscellaneous tips and information about the method, links to articles they have published about the method, and copies of the text files used in their *APSR* article (to use for tutorial purposes). We would like to thank Kenneth Benoit in particular for his technical assistance.
[10] The *Wordscores* program treats any string that begins with a letter as a word.

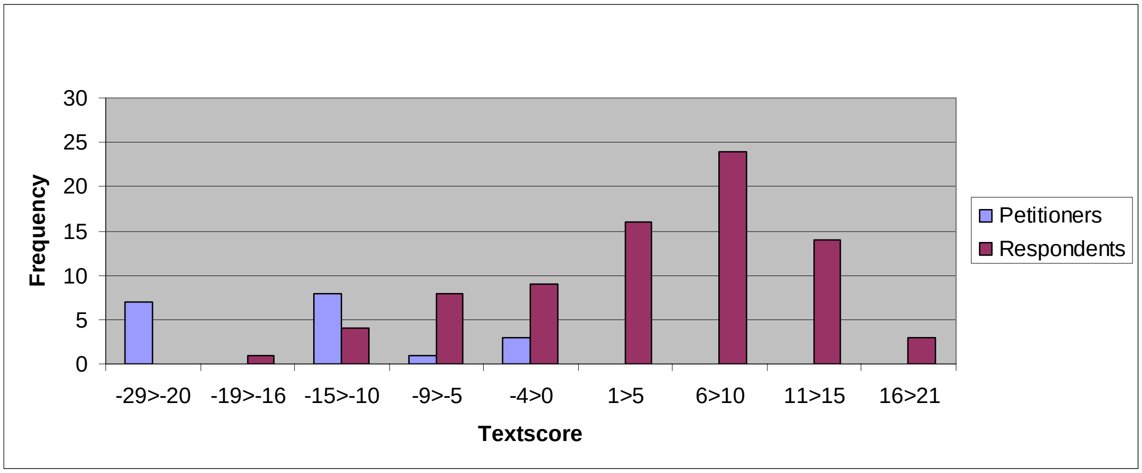
textscore that places it close to D_2 on the set dimension (see Laver et al., 2003, pp. 314-316).

To test the effectiveness of the *Wordscoring* technique for detecting the policy position of legal briefs, we have performed a simple analysis utilizing the principal party and *amicus curiae* briefs from the two *Bollinger* affirmative action cases. To do so, we assume that briefs written by or for petitioners[11] articulate a conservative, anti-affirmative action position, and, conversely, respondent briefs espouse a more liberal, pro-affirmative action position. To verify this, we coded the argument headings of each brief.[12] As expected, although all parties utilize a rich diversity of arguments, each brief can meaningfully be classified based upon its declared support for either respondents or petitioners.[13] In all, 87 distinct arguments were identified. Of these arguments, *amici* for the two positions explicitly used only two in common: that strict scrutiny is the appropriate standard of review and that Justice Powell's opinion in *Bakke* has precedential value. The former argument was explicitly offered in the argument headings of four of the 79 respondent *amicus* briefs and in two of 19 petitioner briefs, while the latter was used in the headings of six briefs for respondents and in one for petitioners. The remaining arguments were used exclusively by one side or the other, and were widely dispersed among the participants. Furthermore, as we shall discuss below, advocates for the opposing positions used discernibly distinct language that can be interpreted along ideological lines.

Utilizing the four principal *Grutter-Gratz* briefs as reference texts, we set out to determine whether we can detect the policy positions of *amici* briefs with the *Wordscoring* method (by attempting to classify them as either supporting the petitioner or supporting the respondent). We assigned the two petitioner briefs a value of -10 and the two respondent briefs a value of 10. All *amicus curiae* briefs were treated as virgin texts. Our decision rule is straightforward: any document with a score below zero is hypothesized as espousing an anti-affirmative action position and any *amicus* scoring above zero is considered pro-affirmative action. Knowing nothing about the distribution of virgin texts, one would expect random guessing to produce, at best, a .50 classification accuracy. Knowing the distribution, by contrast, one would do best to guess that each text is a respondent brief, yielding a success rate of 79 out of 98, or a classification accuracy of about 81%. The *Wordscores*-driven detection method, which requires no knowledge about the prior distribution of virgin texts, yielded a success rate of 76 out of 98, or a classification accuracy of about 76%. Unlike with the method of guessing based on knowledge of the underlying distribution of the briefs, the "misses" under this method are more evenly distributed between the two sides. While the guessing method would yield a 100% success rate at identifying respondent *amicus* briefs, it would fail to detect any of the petitioner *amicus* briefs. By contrast, the *Wordscores* method accurately classified 17 of 19 *amici* for petitioners (89% accuracy), and correctly classified 59 of the 79 *amici* for respondents (75% accuracy).

[11] Recall that 19 amici declared support for petitioners, 89 declared for respondents, and 6 were "neutral".
[12] This is the technique used by Spriggs and Wahlbeck (1997) to assess the argument contributions of *amici* during the 1992 Supreme Court term.
[13] Below we discuss important qualitative differences among the words systematically favored by the respective sides.

Figure 2: Distribution of *Wordscores*-Generated Textscores of *Amicus Curiae* Briefs by "True" Position (Declared Support for Either Respondent or Petitioner) Using Principal Party Briefs as Reference (Training) Documents[14]



Interestingly, this particular combination of reference and virgin texts yielded an overall distribution biased towards the petitioner position (shown in Figure 2). Although both positions are clustered together, we see far more respondent briefs below zero than petitioner briefs above zero. Furthermore, the mean score for petitioner briefs, -14.7, is nearly five points "more conservative" than the set reference value (-10), while the respondent briefs' mean score of 4.6 is nearly five points "less liberal" than their respective reference value (10).[15] Both groups of texts had relatively large standard deviations, about 9.6 for the petitioner briefs and about 8.0 for respondents. Petitioner brief textscores ranged from -29 to .3169, while respondent textscores ranged from -15.88 to 20.79.

Based on this particular collection of texts, our chosen decision rule locates too many documents on the petitioner side of this issue. Why might this be? One possibility is that this is a consequence of greater heterogeneity in argumentation by respondents *viz à viz* petitioners. Put another way, as a group, those opposing affirmative action are far more homogenous, or doctrinally similar, in their arguments than is true among those supporting the programs. While this might make intuitive sense, our coding analysis of the argument headings casts doubt on this. As mentioned above, it appears that both sides utilize a diverse set of arguments to support their positions.[16] While it is possible that

[14] N = 98 (19 petitioner and 79 respondent *amicus* briefs). All documents are from the *Gratz* (539 U.S. 244) and *Grutter* (539 U.S. 982) affirmative action cases (2003), and were acquired using *Lexis Nexis.*
[15] We are using the "transformed textscores" generated by the *Wordscores* program, which is why the generated textscores can exceed the set reference text parameters of -10 and 10. See Laver, et al. (2003, p. 316) for more information about these transformed textscores.
[16] The groups writing on behalf of respondents do represent a more diverse cross-section of society than those writing on behalf of petitioners. Included among the respondent *amici* were corporations, former army officers, congress members, university/law school deans and professors, law students, as well as various progressive political and legal advocacy organizations. By contrast, no corporations,

different coding rules would yield a different conclusion, we think another explanation may nevertheless be closer to the mark.

Another possibility is that petitioner briefs are more readily detectable because groups and individuals opposing affirmative action seem to draw from a more homogenous set of words than do those arguing in favor of affirmative action. In this case, the language utilized by those arguing on behalf of petitioners was notably abstract, legalistic, deontic,[17] skeptical, and individualistic. We will discuss this further below, but, for now, suffice it to say that this *lexical consistency across disparate arguments* among those writing on behalf of petitioners appears to be the best explanation we have for why the anti-affirmative action position is easier to recognize. Furthermore, the legalistic nature of the conservative lexicon can also help explain why some respondent briefs were incorrectly identified by the *Wordscoring* technique. This is supported by the textscores generated by the six "neutral" *amicus* briefs. Among these, four were clearly pro-affirmative action, and one was clearly opposed. These were all correctly detected by the *Wordscores* method. The sixth neutral brief, however, written by the Criminal Justice Legal Foundation, appears to articulate a truly neutral position, offering only a highly technical argument that Powell's opinion in *Bakke* lacks precedential value. Despite its genuine neutrality, the textscore was a very low -22.66. We take this as preliminary evidence that this particular use of the *Wordscores* method places legalistic language in the anti-affirmative action position. This is, in part, a function of our using the principal parties' advocacy documents as reference texts. Indeed, the petitioners rely heavily upon legalistic rhetoric, with frequent references to such concepts as "equal protection" and the value of precedent, whereas the respondents focused more on the inherent beneficial social and economic values of the policy. Regardless, these explanations are only tentative absent further experimentation.

5c. Analysis using a Naïve Bayes' Classifier. It is important to note that our decision rule of grouping documents based on their textscores essentially converts the *Wordscoring* technique into a simple classifier. Thus, even with existing techniques, we have demonstrated the value of approaching legal analysis using the text classification framework. In this section, we compare the *Wordscoring* method with a Naïve Bayes' classifier (Lewis 1992, 1998; Larkey and Croft, 1996; Koller and Sahami 1997; Joachims 1998), a commonly used machine learning algorithm for text classification.

In truth, a Naïve Bayes' classifier is not all that different from the *Wordscoring* method; both are based on the relative frequencies of particular words across documents of different types (e.g., petitioner vs. respondent). Naïve Bayes', however, has the advantage that it places the frequency calculations on the solid theoretical foundation of the laws of probability. Let us suppose that based on the observation of a particular word w_1 from document D_1, one had to guess whether D_1 is a petitioner brief (P) or a

representatives from the military or Congress, or university/law school deans wrote on behalf of the petitioners. Nearly all petitioner *amici* were conservative legal/political advocacy organizations or affiliated with either the Florida Governor or U.S. President Bush administrations. Nevertheless, it does not appear that the respondents used a (proportionally) greater number of different arguments than the petitioners.

[17] That is, concerned with agent-centered moral or legal duties or prohibitions.

respondent brief (R). The probability that D_1 supports, say, the respondents can be derived by application of Bayes' Theorem (Borel 1965):

$$P(R \mid w_1) = \frac{P(w_1 \mid R)P(R)}{P(w_1)}$$

The probability that a particular document containing the word w_1 is a respondent brief is equal to the product of the probability that a respondent brief contains the word in question and the probability that a randomly chosen brief is a respondent brief, divided by the probability of seeing w_1 (in any brief). In order to classify the document, one should simply choose the class with the highest probability. We get $P(w_1|R)$ from the training examples: like the *Wordscoring* method, classification decisions are ultimately based on how frequently a word appears in each type of document. A Naïve Bayes' classifier is based on the same intuition, except that it aggregates evidence from many individual words based on the assumption that all words are conditionally independent. This independence assumption gives the algorithm it's "naïve" label, because it is often violated in real-world texts; for example, the phrase "death penalty" occurs much more frequently than the independent occurrence of the individual words "death" and "penalty."

What words should be employed as features in our classification task? Although a Naïve Bayes classifier can handle an arbitrarily large number of features, many words are not discriminative. For example, both respondents and petitioners use the word "education," so the presence of that particular word does not provide a good indication of the document's proper label. The most useful words for our classification task are ones that are used frequently by one side, but not by the other. This idea can be formally captured in terms of Information Gain, an information-theoretic measure defined in terms of the decrease in Shannon's entropy based on a particular observation (Manning and Schütze 1999). Using Information Gain measure, we selected the top 50 most discriminative words as the features for our classification problem.

We also recognize that words are inherently of different quality, independent of their discriminative power with respect to particular classes. All things being equal, more importance should be assigned to a word that appears frequently in a particular document; the more often a word occurs, the more likely the document is "about" the concept evoked by that word. On the other hand, however, words that appear in many documents are not useful for capturing textual content (in the extreme case, consider stopwords such as "the," "of," "a," and the like). These insights are captured using *tf.idf* term weighting, which is commonly used in information retrieval tasks (Salton 1975; Robertson 2004). With this method, a feature (i.e., word) is assigned a weight equal to the product of its term frequency (*tf*) and inverse document frequency (*idf*):

$$w_{i,j} = tf_{i,j} \times idf_i$$
$$tf_{i,j} = c_{i,j} / t_j$$
$$idf_i = log(N/d_i)$$

where $c_{i,j}$ = number of occurrences of term i in document j
t_j = number of total words in document j
N = number of documents in the collection
d_i = number of documents where term i occurs

In our first experiment with Naïve Bayes', we employed the four principal *Grutter-Gratz* briefs as training examples, using the top 50 most discriminative words as features and weighting them according to the *tf.idf* scheme. Testing was performed on the *amici* briefs. This classifier correctly labeled 75 out of 98 briefs, which corresponds to 76% classification accuracy.

Why did the use of a more sophisticated machine learning algorithm result in essentially the same performance as the *Wordscoring* method? As previously mentioned, the techniques share similar intuitions, but vary in their details. The deeper explanation, however, points to the nature of the task itself: it is difficult to correctly classify the *amici* briefs because they are significantly different in terms of linguistic content from the original litigants' briefs, to the extent that words used in those original briefs do not adequately capture all the arguments forwarded by either side. This confirms the findings of Spriggs and Wahlbeck study (1997, 372), who reported that a minority (although a sizeable minority, 25-35%) of all *amici* exclusively reiterate arguments presented by the principal parties, while about a quarter of them exclusively add information to the process. Based on our manual analysis of the arguments, this is certainly the case: *amici* briefs employ many arguments that are not found in our training examples, the original *Grutter-Gratz* briefs. Of course, this is not surprising: *amicus* briefs are expected to provide alternative viewpoints on the case in question (see e.g., Barker 1967; Caldeira and Wright 1998; Epstein and Kobylka 1992; Epstein 1993; and cf. Spriggs and Wahlbeck 1997).

To test this hypothesis, we explored the possibility of using the *amicus* briefs as training data to predict the policy position of other *amicus* briefs. This is accomplished with a technique called cross-validation. With cross-validation, a single dataset is divided into sections called *folds*. The classifier is trained on some of these folds, and then tested on the rest. By rotating through which folds are selected for training and testing, we can get an accurate picture of performance while maintaining the split between training and test examples. For our second experiment, we employed the same basic setup as in our first experiment, but performed ten-fold cross-validation on the collection of *amici* briefs with a 90/10 split. This means that we divided the briefs into ten equal portions, trained on nine of them, and tested on the other portion; this process was repeated ten times with different partitions of the training/testing examples. Overall, the resulting classifier was able to correctly label 96 of the 98 *amici* briefs, for a classification accuracy of 98%. This result confirms our original hypothesis that the linguistic content of litigants' briefs do not provide an accurate model of the overall arguments used by both sides.

5d. Disaggregating the analyses: what can we learn from "petitioner words" and "respondent words"? The above analyses demonstrate the potential for applying computational methods to automatically infer the policy positions of legal briefs based solely on linguistic content. The 76% accuracy rates of the *Wordscoring* and Naïve Bayes analyses, and 98% accuracy rate of the cross-validation results, are quite impressive for automated processes. Of course, merely identifying the policy positions of *amicus* briefs is of limited utility; most third-parties do, after all, explicitly declare their support for one side or the other. The purpose of our preliminary work, though, is to demonstrate the validity of our methodology on a simple problem where training and test examples are readily available. In principle, the same set of analytical tools could be employed by any researcher to explore any large collection of interesting texts. Legal scholars may especially find classification techniques useful as a "first-cut" analysis of large document sets such as memos, transcripts, and unidentified briefs,[18] with truly *a priori* unknown positions. Indeed, any conceivable practical application of computational techniques would most likely need to be treated as a "first-cut;" automated systems can never replace the insights that scholars bring to bear. As with a Web search, we are ever-cognizant that results will rarely be 100% accurate, but we can nevertheless use the tool for its speed and efficacy at providing a relatively high percentage of relevant texts to manually analyze and verify.

However, computational methods can also serve a deeper theoretical purpose for judicial scholars, beyond mere text classification. That is, by asking *why* automated detection works, we encounter theoretically salient implications about the nature and role of language and ideology in the behavior of legal actors. As mentioned earlier, despite the fact that we identified a diverse distribution of arguments within and across the groups declaring support for each side in the *Bollinger* cases, groups identifying with either the petitioner or respondent use similar words as other groups sharing the same position. Indeed, if this were not so, automated classification techniques could not work. As also mentioned above, the words that are used at a relatively high frequency by one side and not the other are identified through the measure of Information Gain (IG). Interestingly, it appears that the words with highest IG do not only indicate different arguments by petitioners and respondents, but also reveal qualitatively different ways of conceptualizing the issue of affirmative action.

Table 1 lists the highest IG words indicative of respondent briefs and Table 2 lists the highest IG petitioner words.[19] By using the content analysis tools of Provalis' *QDA Miner* (v1.1) and *Wordstat* (v4.0.20), we observed these words in context. This revealed several qualitative differences in style of argumentation offered by the two sides. In general, respondent groups used language associated with an emphasis on the *impact* of affirmative action polices, while petitioner words indicate concern over legal and administrative *procedure*. High IG respondent words are associated with concern about the concrete consequences of affirmative action policies on the (domestic and "global") "market" economy (and business interests), the "recruitment" and "training" of next generation "leaders" in the "labor" force (and military), and the achievement of substantial "opportunities" for all citizens, including the "poor," racial minorities, historically oppressed groups, and those from "underdeveloped"—especially "urban"—"areas." Furthermore, an equality-of-opportunity, rather than strict egalitarian, conception of justice seems to inform respondents' arguments about the impact of affirmative action policies. The respondent's emphasis upon impact requires, of course, at least a minimal epistemic optimism in our ability to understand social and economic causal processes.

---

[18] As mentioned above, the *Wordscores* method accurately identified the true (non-neutral) positions of five out of the six *amicus* briefs claiming to be "neutral." (And the sixth probably was truly neutral and thus non-identifiable).

[19] Minimal discretion was used in selecting these particular words. They all were drawn from among the 150 over-all highest IG words. Most of the excluded strings were numbers and names, categories of words that were probably indicative of favored citations by each side. Of course, distinct citation habits are nontrivial tendencies, but not particularly helpful for the present analysis.

Respondents argue that “projected” consequences of alternative courses of action point to the “vital” need for affirmative action policies. Respondents’ teleological orientation also incorporates arguments about the ability (and implied right) of institutions (states, acting through universities) to “shape” outcomes. The focus, nonetheless, is upon the social, economic, and national consequences of affirmative action policies and not upon the legal and administrative procedures used to create and implement those policies, or their individualized effects.

In stark contrast, high IG petitioner words reflect an abstract focus on legal justification/procedure, epistemology, deontological duties/prohibitions, and individualized (either meritocratic or libertarian) justice. The proceduralist words take many forms, but they all are used in the context of arguments claiming that affirmative action procedures are somehow illegitimate. Some argue that the policies “unjustifiably” show “preferential” treatment towards “beneficiaries” based on “vague,” “indefinite,” “random,” “unreliable,” and/or “amorphous” “categories” such as “skin” color. The procedures are “forbidden,” many claim, because they “reject” and “violate” the “concept” of equal “protection” as guaranteed by the constitution. To the extent that these words are associated with the consequences of affirmative action, the relation is based either on skepticism of respondents’ claims about the “benign” impact of the policies or assertions of the perverse or “dangerous” unintended consequences of the policies. Many petitioners doubt that diversity is a compelling state interest as “purported,” or that it actually delivers many of its “alleged” benefits. In fact, some argue that it actually unduly “burdens” the “innocent” while actually “stigmatizing” its “supposed” “beneficiaries.” Another common claim is that since the history of past discrimination was unjust, it is “dangerous” to use any criteria for admissions today other than “merit” or a “code” somehow “logically” “justified.” While, again, these (and related) words are combined in various ways to make several distinct arguments, a common thread uniting petitioners appears to be a heavy reliance upon words that connote proceduralism, legalism, skepticism, and individualism.

6. Conclusions

The primary aim of this study has been to explore the utility of employing computational techniques to study a relatively large collection of legal briefs. The promise of these methods is that students of the Court will be able to increase substantially the quantity of legal texts used in their studies, and to employ quantitative techniques for comparing and analyzing texts.

Our preliminary analysis does indicate that quantitative approaches in general, and our text classification framework in particular, hold considerable promise. Textscores generated for the *amici* briefs submitted for the opposing sides of the affirmative action debate make good intuitive sense, and the wordscores help to isolate the primary issues that the two sides emphasized repeatedly. Moreover, the cross-validation experiment with our Naïve Bayes classifier not only correctly identified the issue positions of nearly all the briefs in our sample, but the associated Information Gain measure yielded significant insight into the arguments presented to the Court by the advocates on either side of the affirmative action divide. We are confident that these methods can be applied to any competing set of adversaries to analyze arguments presented in supporting documents. But there are certainly more questions than we have addressed here: Do different types of groups (e.g., membership-based, commercial, social advocacy, governments, and so on) participating as *amicus curiae* make (discernibly) distinct types of arguments? Are there important differences among the parties on each side—i.e., are there significant characteristics among those falling above and below the means of each distribution? The technique also holds potential to determine consistency of argument over a period of years in a large number of case iterations. For example, do the ACLU and Pacific Legal Foundation, long-term players in the affirmative action controversy (as well as a range of others), present consistent positions over time, or even across issue fields?

Our analysis of the 98 *amicus* briefs (19 petitioner and 79 respondent *amici*), which were written by groups with diverse interests and perspectives and which presented very different arguments, nevertheless revealed two general *styles* of argumentation by the opposing sides. What might this suggest about the nature of legal rhetoric and activism? At this stage, we hesitate to offer more than hypotheses for further inquiry. The primary puzzle at this point is why one group adopted a distinctly abstract and proceduralist style of argumentation while the other argued with an emphasis on concrete consequences. One hypothesis would be that these styles of argumentation reflect the *context-independent* cognitive processes of their authors; that groups on either side are united by ideological predispositions that are revealed through their common styles of legal argumentation. To put it simply, petitioner-supporters used proceduralist arguments because they are proceduralist conservatives. Accordingly, law is viewed as essentially conservative, with a past orientation, and the courts provide a proper forum where parties can seek individualized remedies for wrongs accruing from implementation of policies enacted elsewhere. The courts are not the proper forum to forge or craft social policies. Respondent-supporters used consequentialist arguments because they begin with a very different view of the role of law and courts, believing that the judicial forum can be utilized to promote and refine social policy. Another hypothesis would predict that one's argumentative style is a function of the *strategic context* within which one argues. That is, if proceduralist arguments are expected to be the most persuasive instruments for achieving one's preferred outcome, then one will adopt proceduralist arguments. If one believes an emphasis on consequences will prevail, then one will choose that rhetorical course. Rhetorical strategy might also be used to explain the abstract (rather than concrete) orientation of petitioners: if the proscription of affirmative action will exacerbate social inequalities (along both class and racial lines) and weaken the strength of the nation economically and militarily, then a de-emphasis of concrete consequences in favor of an abstract axiomatic style of argumentation may indeed be one's dominant strategy.

The *cognitive (context independent)* vs. *strategic-context* explanations represent very different perspectives on the nature of legal argumentation. The crucial difference lies in the degree of *autonomous rhetorical discretion* attributed to authors. If an *amicus* brief author is able to make a rigid separation between the beliefs underlying her political preferences and the style of argumentation she adopts to advance those preferences, we would say she has a great deal of autonomous rhetorical discretion. On the other hand, if an author's arguments are to a significant degree influenced by her underlying political preferences, then she has much less rhetorical discretion. To the extent that underlying preferences are a product of ideological dispositions, and those same dispositions are intertwined with attitudes towards the proper role of government institutions in society, then we would expect legal arguments by authors with low rhetorical discretion to reveal underlying ideological dispositions. It may be that legal actors vary in their levels of rhetorical discretion, meaning the correspondence between legal arguments and underlying ideologies varies across authors. For instance, it may be the case that U.S. Supreme Court justices are more skilled at separating their ideological beliefs and motives from the jurisprudential language they use to advance their preferences than *amicus* authors. They also write for different purposes and audiences. A justice writes to explain a decision, to provide guidance for subsequent action, as well as for other decisional-strategic reasons. An *amicus* brief author writes to advocate a position. Repeat-player *amici* may also have more highly developed levels of rhetorical discretion than less seasoned *amici*. The best way to examine these questions will be to generate more lists of high IG words by different sides and legal actors within different contexts and assess the patterns that emerge. We may indeed find that both the *context-independent* and *strategic-context* hypotheses have explanatory power, but that their relative merit relies upon the role and experience of different agents.

We also believe that computational methods can provide us with leverage on the thinking of the Justices. As with the advocates, we feel these procedures offer great potential for determining consistency of opinion offered by individual members of the Court on this set of issues as well as others, and for gaining additional insights into perspectives of the justices. In addition, the tools described in this paper provide a means for bringing innovative and fresh approaches to the analysis of judicial decision-making and opinion writing. We believe that these techniques offer an excellent opportunity to explore the possible influences of litigants and *amici* on judicial opinions, ultimately leading to a deeper understanding of the complex roles that key actors play within the judicial system .

Table 1: Pro-Affirmative Action Words:
Frequency and Interpretation of High "Information Gain" Words Associated with *Amicus Curiae* Briefs Written on Behalf of *Gratz and Grutter* Respondents[20]

Word	Respondent Frequency	Petitioner Frequency (x 5)[21]	Ratio Resp : Pet	Rhetorical Tone/Style/Emphasis
MARKET*	108	0		Impact: Econ
WORKFORC*	103	0		Impact: Econ
MANAG*	104	5	20.80	Impact: Econ / Social
FINANCI*	104	10	10.40	Impact: Econ
VITAL*	104	10	10.40	Urgency / Salience
RECRUIT*	133	15	8.87	Impact: Econ / Social / Milit
TRAIN*	211	25	8.44	Impact: Econ / Social / Milit
PRIVAT*	246	35	7.03	Impact: Econ / Soc
LEADER*	317	50	6.34	Impact: Econ / Social / Milit
RESOURC*	112	20	5.60	Impact: Econ / Social / Milit
EXCHANG*	109	20	5.45	Impact: Econ
WORLD*	184	35	5.26	Impact: Econ / Global
IMPACT*	224	75	2.99	Impact
POOR*	133	45	2.96	Impact: Econ / Social Justice
AREA*	192	70	2.74	Impact: Social Justice
SHAPE*	153	60	2.55	Impact
DEVELOP	271	135	2.01	Impact: Social Justice
INFORM*	170	85	2.00	Impact
ECONOMI*	204	110	1.85	Impact: Econ
POSIT*	314	185	1.70	Impact
OPPORTUN*	418	330	1.27	Impact: Econ / Social Justice
ALUMNI*	96	20	4.80	Procedure
EXPOSUR*	96	35	2.74	Impact
INFLUENC*	95	15	6.33	Impact
HEREINAFT*	91	0		Legalistic
DRAW*	88	95	0.93	Procedure
GLOBAL*	84	0		Impact: Econ / Global
PROJECT*	81	5	16.20	Expected Future Impact
LABOR*	79	0		Impact: Econ / Social
URBAN*	66	5	13.20	Impact: Society
MULTIPL*	61	15	4.07	Procedure

[20] High IG words were identified using the Naïve-Bayes classifier described in section 5c. Only term frequencies are reported here, although, as discussed above, each word is actually weighted according to the product of its term frequency and inverse document frequency. The reported frequencies were calculated and words interpreted, in part, by utilizing Provalis' *QDA Miner* (v1.1) and *Wordstat* (v4.0.20).
[21] Due to the considerably greater number of respondent *amicus* briefs, respondents used roughly 5 times as many words as petitioners. Thus, the total number of petitioner words is multiplied by 5 in Table 1 and Table 2 to allow for meaningful comparisons.

Table 2: Anti-Affirmative Action Words:
Frequency and Interpretation of High "Information Gain" Words Associated with *Amicus Curiae* Briefs Written on Behalf of *Gratz and Grutter* Petitioners

Word	Respondent Frequency	Petitioner Frequency (x 5)	Ratio Pet:Resp	Rhetorical Tone/Style/Emphasis
VAGU*	1	55	55.00	Procedure / Skepticism
INNOC*	4	55	13.75	Intent
INDEFINIT*	2	55	27.50	Procedure / Skepticism
RANDOM*	13	50	3.85	Procedure
CODE	19	45	2.37	Procedure
BENIGN	7	120	17.14	Intent
PREFERENTI*	14	210	15.00	Procedure
PREFER*	175	2175	12.43	Procedure
CLASSIF*	7	65	9.29	Procedure / Deontic
UNREL*	7	65	9.29	Procedure / Skepticism
AMORPH*	12	10	9.17	Procedure / Skepticism
DANGER*	15	115	7.67	Dangerous Consequences
JUSTIF*	50	360	7.20	Procedure / Legalistic
SUPPOS*	16	105	6.56	Skepticism
CLASSIFI*	147	890	6.05	Procedure / Deontic
BENEFICIARI*	16	95	5.94	Impact (Individualized)
PURPORT*	32	180	5.63	Skepticism
SKIN*	34	180	5.29	Procedure
LOGIC*	27	130	4.81	Legalistic / Axiomatic
FORBID*	25	115	4.60	Legalistic / Deontic
CATEGOR*	59	250	4.24	Procedure / Deontic
STIGMA*	20	80	4.00	Perverse Consequences
BURDEN*	73	280	3.84	Procedure / Legalistic
FAVOR	113	415	3.67	Procedure / Deontic
CLAIM*	123	405	3.29	Procedure / Skepticism
DEFIN*	117	340	2.91	Procedure
CONCEPT*	58	160	2.76	Procedure / Deontic
PLURAL*	82	215	2.62	Procedure / Epistemic
VIOLAT*	104	265	2.55	Social
REJECT*	173	435	2.51	Legalistic / Deontic
PROTECT*	372	885	2.38	Procedure / Legalistic
MERIT*	104	230	2.21	Procedure / Deontic
ALLEG*	31	65	2.10	Procedure / Justice

References

Barker, Lucius. 1967. "Third Parties in Litigation: A Systematic View." 29 *Journal of Politics* 41-69.

Bergara, M., B. Richman, et al. 2003. "Modeling Supreme Court Strategic Decision Making: The Congressional Constraint." 28 *Legislative Studies Quarterly* 247-280.

Borel, E. 1965. Elements of the Theory of Probability. Englewood Cliffs, NJ: Prentice Hall.

Brenner, S. and M. Stier. 1996. "Retesting Segal and Spaeth's Stare Decisis Model." 40 *American Journal of Political Science* 1036-1048.

Brill, E. and Mooney, R. 1997. "An Overview of Empirical Natural Language Processing." AI Magazine, 18(4):13-24.

Brisbin, R. A., Jr. 1996. "Slaying the Dragon: Segal, Spaeth and the Function of Law in Supreme Court Decision Making." 40 *American Journal of Political Science* 1004-1017.

Caldeira, G. A., and J. R. Wright. 1988. "Organized Interests and Agenda Setting in the U.S. Supreme Court." 82 *American Political Science Review* 1109-1127.

Caldeira, G. A., and J. R. Wright. 1990. "Amici Curiae before the Supreme Court: Who Participates, When, and How Much?" 52 *The Journal of Politics* 782-806.

Caldeira, Gregory A. and Christopher J.W. Zorn. 1998. "Of Time and Consensual Norms in the Supreme Court." 42 *American Journal of Political Science* 874-902.

Carmines, Edward, and Richard Zeller. 1979. *Reliability and Validity Assessment.* Beverly Hills, CA: Sage.

Danelski, David J. 1966. "Values as Variables in Judicial Decision-Making: Notes Toward a Theory." 19 *Vanderbilt Law Review* 721-40.

Epstein, Lee. 1993. "Interest Group Litigation During the Rehnquist Court Era," 9 *Journal of Law and Politics* 639-717.

Epstein, Lee, and Joseph F. Kobylka. 1992. *The Supreme Court and Legal Change: Abortion and the Death Penalty*. Chapel Hill: University of North Carolina Press.

Epstein, Lee and Carol Mershon. 1996. "Measuring Political Preferences." 40 *American Journal of Political Science* 261-294.

Epstein, Lee, and Jack Knight. 1998. *The Choices Justices Make*. Washington, DC: CQ Press.

Epstein, Lee and Jack Knight. 1999. "Mapping Out the Strategic Terrain: The Informational Role of Amici Curiae," in Cornell W. Clayton and Howard Gillman. Supreme Court Decision-Making: New Institutional Approaches. Chicago: University of Chicago Press, 215-235.

Epstein, Lee and Jack Knight. 2000. "Toward a Strategic Revolution in Judicial Politics: A Look Back, a Look Ahead." 53 *Political Research Quarterly* 625-661.

Epstein, Lee, Jeffrey A. Segal, and Harold J. Spaeth. 2001. "The Norm of Consensus on the U.S. Supreme Court." 45 *American Journal of Political Science* 362-377.

Epstein, Lee, Jeffrey A. Segal, Harold J. Spaeth, and Thomas G. Walker. 2003. *The Supreme Court Compendium: Data, Decisions, and Developments* (3rd ed). Washington, DC: CQ Press.

Eskridge Jr., W. N. 1991. "Reneging on History?: Playing the Court/Congress/President Civil Rights Game." 79 *California Law Review* 613.

Foster, Donald W. 1989. *"Elegy" by W.S.: A Study in Attribution*. Cranbury, NJ: Associated University Presses.

Foster, Donald W. 2000. *Author Unknown: On The Trail of Anonymous*. New York: Henry Holt.

Frakes, W. and Baeza-Yates, R. 1992. Information Retrieval: Data Structures and Algorithms. New York: Prentice Hall.

Gely, R. and P. T. Spiller. 1990. "A Rational Choice Theory of Supreme Court Decision Making with Applications to the State Farm and Grove City Cases." 6 *Journal of Law, Politics, and Organization* 263.

Gibson, James L. 1997. United States Supreme Court Judicial Database, Phase II: 1953–1993 [codebook] (Study 6987). ICPSR version. Houston, TX: University of Houston [producer], 1996. Inter-University Consortium for Political and Social Research, Ann Arbor, MI [distributor].

Hansford, Thomas G. 2004. "Lobbying Strategies, Venue Selection, and Organized Interest Involvement at the U.S. Supreme Court." 32 *American Politics Research* 170-197.

Hoover, Kenneth R., John Miles, Vernon Johnson, and Sara Weir. (2001). *Ideology and Political Life, 3rd Edition,* Wadsworth Publications.

Howard, J. W., Jr. 1968. "On the Fluidity of Judicial Choice." 62 *The American Political Science Review* 43-56.

Joachims, T. 1998. "Text Categorization With Support Vector Machines: Learning With Many Relevant Features. Proceedings of the 10th European Conference on Machine Learning.

Johnson, T. R. 2001. "Information, Oral Arguments, and Supreme Court Decision Making." 29 *American Politics Research* 331-351.

Johnson, T. R. 2003. "The Supreme Court, the Solicitor General, and the Separation of Powers." 31 *American Politics Research* 426-451.

Johnson, T. R. 2004. *Oral Arguments and Decision Making on the United States Supreme Court*. Albany: State University of New York Press.

Kenny, Anthony. 1986. *A Stylometric Study of the New Testament.* Oxford: Oxford University Press.

Knight, J. and L. Epstein. 1996. “The Norm of Stare Decisis.” 40 *American Journal of Political Science* 1018-1035.

Knight, K. 1999. “Mining Online Text.” Communications of the ACM. 42(11):487-494.

Koller, D. and Sahami, M. 1997. “Hierarchically Classifying Documents Using Very Few Words.” Proceedings of the14th International Conference on Machine Learning (ICML 1997).

Larkey, L. and Croft, W. 1996. “Combining Classifiers in Text Categorization.” Proceedings of the 19th ACM International Conference on Research and Development in Information Retrieval (SIGIR 1996).

Larsen, Wayne A., Alvin C. Rancher, and Tim Layton. 1980. “Who Wrote the Book of Mormon? An Analysis of Wordprints.” 20 *BYU Studies* 225-251.

Laver, Michael, Kenneth Benoit, and John Garry. 2003. “Extracting Policy Positions from Political Texts Using Words as Data.” 97 *American Political Science Review* 311-331.

Levi, Edward H. 1949. *An Introduction to Legal Reasoning*. Chicago: University of Chicago Press.

Lewis, D. 1992. “Representation and Learning in Information Retrieval.” Ph.D. thesis, Department of Computer Science, University of Massachusetts.

Lewis, D. 1998. “Naïve (Bayes) at Forty: The Independence Assumption in Information Retrieval.” Proceedings of the 10th European Conference on Machine Learning (ECML 1998), 1998.

Maltzman, F., J. F. Spriggs, et al. 2000. *Crafting Law on the Supreme Court : The Collegial Game*. Cambridge, UK ; New York: Cambridge University Press.

Maltzman, F. and P. J. Wahlbeck. 1996. “May It Please the Chief? Opinion Assignments in the Rehnquist Court.” 40 *American Journal of Political Science* 421-443.

Manning, C. and Schütze, H. 1999. Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press.

Markoff, John, and Edward Wyatt. 2004. “Google Is Adding Major Libraries to Its Database.” *The New York Times*. December 14, 2004, p.1A.

Martindale, Colin, and Dean McKenzie. 1995. “On the Utility of Content Analysis in Author Attribution: The Federalist.” 29 *Computers and the Humanities* 259-270.

McIntosh, Wayne V., and Cynthia L. Cates. 1997. *Judicial Entrepreneurship: The Role of the Judge in the Marketplace of Ideas*. Westport, CT: Greenwood.

Mitchell, T. 1996. Machine Learning. New York: McGraw Hill.

Morton, Andrew Q., and James McLeman. 1980. *James: The Genesis of John.* Edinburgh: St. Andrews Press.

Morton, Andrew Q. 1993. *The Authorship of the New Testament Epistles.* University of Glasgow: Department of Computing Science.

Mosteller, Frederick and Wallace, David L. 1964. *Applied Bayesian and Classical Inference: The Case of the Federalist Papers*. Addison-Wesley: Reading, MA.

Murphy, Walter F. 1964. *Elements of Judicial Strategy.* Chicago: University of Chicago Press.

O'Brien, David M. 1999. "Institutional Norms and Supreme Court Opinions: On Reconsidering the Rise of Individual Opinions," in Cornell W. Clayton and Howard Gillman. *Supreme Court Decision-Making: New Institutional Approaches.* Chicago: University of Chicago Press, 91-113.

Post, Robert. 2001. "The Supreme Court Opinion as Institutional Practice: Dissent, Legal Scholarship, and Decisionmaking in the Taft Court" 85 *Minnesota Law Review* 1267-1390.

Pritchett, C. Herman. 1941. "Division of Opinion Among Justices of the U.S. Supreme Court, 1939-1941." 35 *American Political Science Review* 890-90.

Pritchett, C. Herman. 1948. *The Roosevelt Court: A Study in Judicial Politics and Values 1937-1947*. New York: Macmillan.

Pritchett, C. Herman. 1954. *Civil Liberties and the Vinson Court.* Chicago: University of Chicago Press.

Raghavan, V. and Wong, S. 1986. "A Critical Analysis of Vector Space Model in Information Retrieval." Journal of the American Society for Information Science, 37:279-287.

Robertson, S. 2004. "Understanding Inverse Document Frequency: On Theoretical Arguments for IDF." Journal of Documentation 60:503-520.

Rohde, David W. 1972. "Policy Goals and Opinion Coalitions in the Supreme Court." 16 *Midwest Journal of Political Science* 208-224.

Rohde, David W., and Harold J. Spaeth. 1976. *Supreme Court Decision Making.* San Francisco: W.H. Freeman.

Salton, G. 1975. "A Vector Space Model for Information Retrieval." Communications of the ACM, 18(11), 613-620.

Salton, G. 1989. Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer. Reading, Massachusetts: Addison-Wesley.

Schubert, Glendon A. 1959. *Quantitative Analysis of Judicial Behavior*. Glencoe, IL: Free Press.

Schubert, Glendon. 1965. *The Judicial Mind: The Attitudes and Ideologies of Supreme Court Justices, 1946-1963.* Evanston, IL: Northwestern University Press.

Schubert, Glendon. 1974. *The Judicial Mind Revisited: Psychometric Analysis of Supreme Court Ideology.* New York: Oxford University Press.

Segal, Jeffrey A. 1991. "Courts, Legislatures, and Executives." In *The American Courts,* John B. Gates and Charles A. Johnson (eds.). Washington, DC: CQ Press.

Segal, Jeffrey A. 1997. "Separation-of-Powers Games in the Positive Theory of Congress and Courts." 91 *American Political Science Review* 28-44.

Segal, Jeffrey A., Lee Epstein, Charles M. Cameron and Harold J. Spaeth. 1995. "Ideological Values and the Votes of U.S. Supreme Court Justices Revisited." 57 *Journal of Politics* 812-23.

Segal, Jeffrey A. and Albert D. Cover. 1989. "Ideological Values and the Votes of U.S. Supreme Court Justices." 83 *American Political Science Review* 557-65.

Segal, Jeffrey A., and Harold J. Spaeth. 1993. *The Supreme Court and the Attitudinal Model.* New York: Cambridge University Press.

Segal, Jeffrey A., and Harold J. Spaeth. 1996. "The Influence of Stare Decisis on the Votes of United States Supreme Court Justices." 40 *American Journal of Political Science* 971-1003.

Segal, Jeffrey A., and Harold J. Spaeth. 1996. "Norms, Dragons, and Stare Decisis: A Response." 40 *American Journal of Political Science* 1064-1082.

Segal, Jeffrey A., and Harold J. Spaeth. 2002. *The Supreme Court and the Attitudinal Model Revisited.* Cambridge: Cambridge University Press.

Shapiro, Martin. 1968. *The Supreme Court and Administrative Agencies.* New York: Free Press.

Songer, D. R. and S. A. Lindquist. 1996. "Not the Whole Story: The Impact of Justices' Values on Supreme Court Decision Making." 40 *American Journal of Political Science* 1049-1063.

Songer, Donald R., and Reginald S. Sheehan. 1993. "Interest Group Success in the Courts: Amicus Participation in the Supreme Court." 46 *Political Research Quarterly* 339-354.

Spaeth, H. J. and J. A. Segal. 1999. *Majority Rule Or Minority Will : Adherence To Precedent On The U.S. Supreme Court.* Cambridge: Cambridge University Press.

Spaeth, H. J. and J. A. Segal. 2001. *Majority Rule, Or, Minority Will : Adherence To Precedent On The U.S. Supreme Court*. Cambridge: Cambridge University Press.

Spiller, P. T. and E. H. Tiller. 1996. "Invitations To Override: Congressional Reversals Of Supreme Court Decisions." 16 *International Review of Law and Economics* 503-521.

Spriggs, J. F., F. Maltzman, et al. 1999. “Bargaining on the U.S. Supreme Court: Justices’ Responses to Majority Opinion Drafts.” 61 *Journal of Politics* 485-506.

Spriggs, James F., II, and Paul J. Wahlbeck. 1997. “Amicus Curiae and the Role of Information at the Supreme Court.” 50 *Political Research Quarterly* 365-386.

Sunstein, Cass. 1999. *One Case at a Time: Judicial Minimalism on the Supreme Court.* Cambridge, MA: Harvard University Press.

Tate, C. Neal and Roger Handberg. 1991. “Time-Binding and Theory-Building in Personal Attribute Models of Supreme Court Voting Behavior, 1916-88.” 35 *American Journal of Political Science* 460-480.

Tumer, K. and Ghosh, J. 1996. “Error Correlation and Error Reduction in Ensemble Classifiers.” Connection Science, 8(3-4):385–403.

Ulmer, S. Sidney. 1986. “Are Social Background Models Time-Bound?” 80 *American Political Science Review* 957-967.

Wahlbeck, P. J., J. F. Spriggs, et al. 1998. “Marshalling the Court: Bargaining and Accommodation on the United States Supreme Court.” 42 *American Journal of Political Science* 294-315.

This Federal Judicial Center publication was undertaken in furtherance of the Centerâ€™s statutory mission to develop and conduct education programs for judicial branch employees. The views expressed are those of the authors and not necessarily those of the Federal Judicial Center. i. ii.Â But many Americans also voiced concern over the potential dan-ger of the federal court system authorized by Article III. By one count, 19 of the 103 amendments proposed by the state ratifying conventions called for changes in Article III.2 Indeed, Anti-Federalists sought limits on Article III for much the same reason they sought a bill of rights (especially those protections relating to judicial procedures): They feared that courtsâ€"especially courts of the new and powerful national governmentâ€"could.