Judicial Ghostwriting: Authorship on the U.S. Supreme Court

Jeffrey S. Rosenthal[◊] and Albert H. Yoon[♠]

August 2010

Abstract

Supreme Court justices, unlike the President or members of Congress, perform their work with relatively little staffing. Each justice processes the docket, hears cases, and writes opinions with the assistance of only their law clerks. The relationship between justices and their clerks is of intense interest to legal scholars and the public, but remains largely unknown. This paper analyzes the text of the justices' opinions to better understand judicial authorship. Based on the use of common function words, we find that justices vary in writing style, from which it is possible to accurately distinguish one justice from another. The justices' writing style also informs how clerks influence the opinion-writing process. Current justices, with few exceptions, exhibit significantly higher variability in their writing than their predecessors, both within and across years. Our results strongly suggest that justices are increasingly relying on their clerks to write opinions.

Introduction

"The reason the public thinks so much of the Justices of the Supreme Court is that they are almost the only people in Washington who do their own work."
- Justice Louis D. Brandeis[1]

Imagine a job where, each year, one is required to evaluate over 7000 files, closely evaluate approximately 60-80 cases, and write 7-10 lengthy published documents, all of which will become established law and be scrutinized by countless judges, lawyers, academics, and law students. Add one additional requirement: do this job well past retirement age, into your seventies and eighties. Incredible as it sounds, the above describes the job of a Supreme Court justice.

In contrast to the other federal branches of government,[2] the Court is a small and closed environment. It consists of only nine justices, who are collectively responsible for deciding

◊ Professor, University of Toronto, Department of Statistics.

♠ Associate Professor, University of Toronto Faculty of Law. We would like to thank Ed Cheng, John Goldberg, Pamela Karlan, Helen Levy, David Madigan, Michael Tribelcock, Fred Tung, and Robert Weisberg. Yoon would like to thank the Russell Sage Foundation for their general financial support for this research. All remaining errors are our own.

[1] CHARLES E. WYNZANKI, JR., WHEREAS—A JUDGE'S PREMISES, ESSAYS IN JUDGMENT, ETHICS, AND THE LAW 61 (1976).

which cases to hear, and ultimately decide in the form of written opinions. Unlike the president or Congress, the Court operates with lean staffing. The justices are each responsible for their own writing; their only assistance is their law clerks, for which each is currently allowed four.[3] Within each justice's chambers, law clerks assist in evaluating *certiorari* petitions, preparing for oral argument, and ultimately producing written opinions. The law clerks are recent law graduates, typically among the top of their class from a handful of elite law schools.[4]

Given the demands of the job, it is understandable if the justices sought to ease their work burdens. They have already taken some steps. For example, the Court in 1972 created a *cert* pool in 1972.[5] Whereas previously each justice evaluated the entire docket of cases on their own, participating justices of the cert pool now reviewed petitions collectively, meaning that they shared work across chambers. This division of labor dramatically reduced the number of petitions each chamber had to process. Moreover, in the *cert* process, it is widely accepted that the law clerks now are largely responsible for providing the initial assessment of the cases.[6]

But anecdotes abound that justices have also increasingly delegated the responsibility of writing opinions to their clerks. One former Justice Stevens clerk reported that clerks generate "well over half" of the text in published opinions.[7] A 2006 historical account of Supreme Court clerks stated, "one can safely conclude that no set of sitting Supreme Court justices have delegated as much responsibility to their law clerks as the Rehnquist Court."[8]

The purported degree to which justices rely on clerks varies considerably. Justice Oliver W. Holmes wrote his opinions in longhand, relegating his clerks to primarily non-legal tasks akin to an administrative assistant.[9] Justice Douglas maintained he wrote his own opinions.[10] Justice

---

[2] For example, each member of the House of Representatives is allowed 18 permanent employees, and up to four additional shared or part-time employees. *See* Ida A. Brudnick, *Congressional Salaries and Allowances*, 7 Congressional Research Services (CRS) Publication (October 28, 2009).

[3] *See* TODD C. PEPPERS, COURTIERS OF THE MARBLE PALACE: THE RISE AND INFLUENCE OF THE SUPREME COURT LAW CLERK 195 (2006).

[4] *See id.* at 30 (describing that justices "pick not only from the best law schools but also the best students from these schools").

[5] *See* ARTEMUS WARD AND DAVID L. WEIDEN, *supra*, SORCERERS' APPRENTICES: 100 YEARS OF LAW CLERKS A THE UNITED STATES SUPREME COURT 117 (2006) (describing the creation of the cert pool during the Burger Court in response to the growing caseload).

[6] *See id.* at 136, 142 (2006) (describing how in 1968, each clerk reviewed an average of 844 cert petitions; by 1972, the number was 271).

[7] Sean Donahue, *Behind the Pillars of Justice: Remarks on Law Clerks,* 3 LONG VIEW 81 (1995).

[8] *See* TODD C. PEPPERS, *supra* note / /, at 191.

[9] *See id.* at 58-59 (2006) (describing Justice Holmes relationship with his law clerks).

[10] *See* transcriptions of conversations between Justice William O. Douglas and Professor Walter F. Murphy, Cassette No. 3, Dec. 20, 1961, Princeton University (cited in WARD AND WEIDEN, *supra* note __, at 205).

Stevens is reputed to draft his own opinions, as is Justice Scalia.[11] Other justices, such as Justice Thurgood Marshall, were known to rely more on their clerks.[12] He was not alone: Chief Justice Rehnquist confirmed that his clerks "d[id] the first draft of almost all cases," and in some instances the published decision was "relatively unchanged" from the draft.[13] Justice Blackmun, in the words of one historian, "ceded to his law clerks much greater control over his official work than any of the other 15 justices from the last half-century whose papers are publicly available."[14]

Should we care if justices delegate the opinion-writing process to their clerks? The answer depends on the degree to which this occurs. The import of an opinion – particularly from the Supreme Court – stems less from the identification of the prevailing party than from the reasoning that accompanies it.[15] Even defenders of delegation, such as Chief Justice Rehnquist, cautioned that each "Justice must retain for himself control not merely of the outcome of the case, but of the explanation of the outcome."[16] Delegation, if taken too far, can threaten the integrity of the Court.

Judicial authorship raises important questions about the relative roles of a justice and her clerks. The extent to which this principal-agent relationship advances the justice's interests depends both on the clerk's competence and preferences. Clerks, while typically excellent students from elite law schools, are also usually recent law graduates: this bimodal age composition on the Court – well-seasoned justices and inexperienced law clerks – lacks a middle cohort with work experience in the government or the private sector. Justice Ginsburg noted that while clerks "save us hours upon hours of labor[,] . . . most of them are also young and in need of the seasoning that experiences in life and in law practice afford." This staffing structure stands in stark contrast to the executive and legislature, replete with experienced staff.

The ideological preferences of the justice and her clerks may also diverge. While liberal and conservative justices generally hire like-minded clerks, the clerkship process – in which clerk applicants apply to all nine justices and are expected to accept the first offer – may brings

---

[11] *See* EDWARD LAZARUS, CLOSED CHAMBERS: THE RISE, FALL, AND FUTURE OF THE SUPREME COURT 271 (2005)

[12] *See* Paul J. Wahlbeck et al., *Ghostwriters on the Court? A Stylistic Analysis of U.S. Supreme Court Opinion Drafts*, 30 Am. Pol. Res. 166, 172 (2002); *see also* Mark Tushnet, *Thurgood Marshall and the Brethren*, 80 Geo. L. J. 2109, 2112 (1992) (writing that "Marshall relied more heavily on his law clerks for opinion writing than did the other Justices during the early years of his tenure, but his practices were not wildly out of line with those of the others on the Court").

[13] *See* BERNARD SCHWARTZ, THE ASCENT OF PRAGMATISM: THE BURGER COURT IN ACTION 38 (1990).

[14] *See* David J. Garrow, "The Brains Behind Blackmun" LEGAL AFFAIRS, May/June 2005.

[15] Justice Oliver Wendell Holmes is attributed as saying, "I am always suspicious of an advocate who comes before the Supreme Court saying this is a court of justice; it is a court of law." EUGENE W. HICKOK & GARY L. MCDOWELL, JUSTICE VS. LAW: COURTS AND POLITICS IN AMERICAN SOCIETY (unnumbered page) (1993)

[16] BERNARD SCHWARTZ, DECISION: HOW THE SUPREME COURT DECIDES CASES 52 (1996).

clerks of different ideologies within the same chambers.[17] Thus, even when the justice dictates the broad direction of an opinion – e.g., the prevailing party and the general reasoning – the clerk may still exercise considerable influence.

Any meaningful discussion of delegation by justices to clerks, however, first requires a deeper understanding of judicial authorship. The question of authorship itself is nothing new and dates back to at least Shakespeare.[18] Much of this earlier work is based on close readings of the text,[19] but ultimately subjective impressions of his writing. Scholars have since turned to statistical analysis. Perhaps most famously, Mosteller and Wallace in their seminal 1964 study of unsigned Federalist Papers, concluded that James Madison – not Alexander Hamilton – was the likely author.[20]

Recent scholarship has examined judicial opinions. A study comparing the first drafts and published opinions by Justices Powell and Thurgood Marshall found that Marshall's clerks writing styles more identifiable in the published opinions.[21] In a study of federal appellate judges, legal scholars used judges' citations to their own earlier opinions as a measure of their own writing, finding a modest relationship between the two.[22] While these studies advance our understanding of the judiciary, neither offer a convincing approach to evaluating judicial authorship: Comparing initial drafts with final opinions may reflect more the collaboration *across* clerks than the justices' reliance on them, while self-citations are a weak proxy for a judge writing her own opinions if her own clerks are also more inclined to cite the judge's opinions.

---

[17] *See* Ward and Weiden, *supra* note / /, at 107-08 (describing the allocation of clerks by ideology across the justices' chambers).

[18] *See, e.g.,* JOHN F. MICHELL, WHO WROTE SHAKESPEARE (1999) (providing survey for arguments for alternative authors); SAMUEL SCHOENBAUM, SHAKESPEARE'S LIVES (1993) (same). *See also* James D.A. Boyle, *The Search for an Author: Shakespeare and the Framers*, 37 AM. UNIV. L. REV. 625 (1988).

[19] Thomas Regnier, Comment, *Could Shakespeare Think Like a Lawyer? How Inheritance Law Issues in Hamlet May Shed Light on the Authorship Question*, 57 U. MIAMI L. REV. 377 (2002) (using legal terms

[20] *See* FREDERICK MOSTELLER & DAVID L. WALLACE, INFERENCE AND DISPUTED AUTHORSHIP: THE FEDERALIST (1964). During this same period, statisticians also examined literature. *See* LOUIS ONKO MILIC, A QUANTITATIVE APPROACH TO THE STYLE OF JONATHAN SWIFT (1967); A.Q. MORTON & JAMES MCLEAN, PAUL, THE MAN AND THE MYTH (1966); ALVAR ELLEGARD, A STATISTICAL METHOD FOR DETERMINING AUTHORSHIP: THE JUNIUS LETTERS, 176901772 (1962).

[21] *See* Paul J. Wahlbeck et al., *Ghostwriters on the Court? A Stylistic Analysis of U.S. Supreme Court Opinion Drafts*, 30 Am. Pol. Res. 166 (2002).

[22] *See* Stephen J. Choi and G. Mitu Gulati, *Which Judges Write Their Opinions (And Should We Care)?*, 32 FLA. ST. U. L. REV. 1077, 1111 (2005) (reasoning that judges who write their own opinions are more likely to cite their own opinions).

Our approach differs from these earlier attempts in that we explore judicial authorship based on a comprehensive evaluation of writing style.[23] The central intuition here is that the more participants in the opinion-writing process, the more heterogeneous the writing style of the justice's opinions. A justice who wrote her own opinions would presumptively possess a less variable writing style than a justice who relied heavily on her law clerks.

The Court's institutional design also provides a compelling identification strategy for our hypothesis. Supreme Court clerkships are typically for a single term, running from October through August. A justice who relies more on her clerks to write opinions would likely have a more variable writing style both within and across years than her less-reliant brethren. In addition, historical accounts of the Court suggests that the responsibilities of clerks has grown over time: from *stenographer* from their inception in the late 19th century to *legal assistant* in the 1920s, to *law firm associate*, beginning in the 1950s.[24] If true, we should expect differences over the history of the Court.

In this paper, we analyze the text of majority opinions of all Supreme Court justices. Using a parsimonious model based on the justices' use of common function words (e.g., *the, also, her*), we construct a variability measure for writing style. Justices have variability scores that are, in most instances, distinguishable from one another. Moreover, even justices with comparable (or identical) variability scores exhibit distinct writing styles based on their use of function words. Our analysis then allows us to accurately predict authorship in pairwise comparisons of justices.

Our model also allows us to evaluate variability in writing both across and within justices. We find that recent justices report significantly higher variability scores than their predecessors, supporting the anecdotal evidence that justices on average are relying more on their clerks in the opinion-writing process. Moreover, some justices – most notably swing justices Justices O'Connor and Kennedy – produce variability scores that are higher and also vary considerably from year to year, suggesting an even greater reliance on clerks. We test the validity of our model by analyzing the opinions of two judges – Richard A. Posner and Frank Easterbrook – known to write their own opinions. Posner and Easterbrook reported variability scores that were markedly lower and more consistent than any of the current justices on the Court.

The paper proceeds as follows. Part II describes our statistical methodology: our construction of function words and the chi-squared approach to evaluate variability in writing style. In Part III we briefly describe the data used in our analysis. We report our results in Part IV: the variability scores, both across and within justices, and our ability to use function words to predict authorship. In Part V we discuss the implications of our results, and how statistical textual analysis can advance future research of the Court and legal scholarship more generally. Part VI concludes.

---

[23] Choi and Gulati briefly note that they attempt to analyze opinions through the judges' choice of words, but find the results inconsistent with their priors of certain judges' reputation for writing their own opinions. *See id.* at 1109.

[24] *See*, PEPPERS, *supra* note //, at 38-144 (describing the changing responsibilities of law clerks from 1886 to the present).

Part II: Methodology

Two factors influence what words an author chooses in her writing. One is subject matter, which can be specific to the author's particular topic or more general to her substantive area.[25] The other is writing style: the sentence structure and choice of words, commonly referred to as *syntax* and *diction*. Our focus is on writing style. The challenge is to tailor our analysis so that it discerns differences in writing style and not subject matter.

Our project falls within the broader discipline of stylometry, the statistical analysis of texts. In many notable endeavors, scholars have attempted to discern authorship of Shakespeare plays,[26] the Federalist papers,[27] and Ronald Reagan's radio addresses.[28]

Our analysis uses function words, which makes use of common words – e.g., *all, have, not, than* – whose usage frequencies are largely independent of subject matter.[29] This approach, in the words of one statistician, is as "topic-free in the sense that the relative frequency with which an author uses, for example, 'with,' should be the same regardless of whether the author is describing cooking recipes or the latest news about the oil futures market."[30]

We considered other approaches, such as larger-scale features such as sentence length, paragraph length, or the frequency of multi-word phrases; or smaller-scale features such as frequency of specific punctuation (e.g,. comma, semi-colon) or particular letters. These approaches, however, did not meaningfully improve our results. To situate our analysis within a unified methodology, we present results only from the use of function words.

In constructing our list of function words, we began by looking at Mosteller and Wallace's seminal study of the Federalist Papers, which had a function list of 70 function words. We adopted this list, but eliminated seven function words – *every, my, shall, should, upon, will, you* – which appeared in fewer than 0.001 of Court majority opinions. Table 1 reports our list of function words.

---

[25] For example, Paul Krugman's columns typically focus on the economic consequences of a specific government policy, but its general focus is on economic policy.

[26] K. Burns, *Bayesian Inference in Disputed Authorship A Case Study of Cognitive Errors and a New System for Decision Support*, 176 INFORMATION SCIENCES 1570 (2006); O. Seletsky et al, *The Shakespeare Authorship Question*, unpublished manuscript, Dartmouth College (2007).

[27] *See* FREDERICK MOSTELLER & DAVID L. WALLACE, INFERENCE AND DISPUTED AUTHORSHIP: THE FEDERALIST (1964).

[28] E.M. Airoldi, et al, *Whose Ideas? Whose Words? Authorship of Ronald Reagan's Radio Addresses*, 40 Pol. Sci. & Pol. 501 (2007); E.M. Airoldi, et al. *Who Wrote Ronald Reagan's Radio Addresses?* 1 BAYESIAN ANALYSIS 289 (2006).

[29] For a more technical explanation of our approach, see Jeffrey S. Rosenfeld and Albert H. Yoon, *Detecting Multiple Authorship of United States Supreme Court Legal Decisions Using Function Words*, forthcoming, _ ANN. APPLIED STAT. __ (2010).

[30] D. Madigan, et al, *Author Identification on the Large Scale,* Proceedings of the Classification Society of North America (CSNA) (2005) (manuscript on file with authors, available at __).

Table 1
Stylometry of the Supreme Court
63 Function Words

(1-12) *a, all, also, an, and, any, are, as, at, be, been, but,*
(13-24) *by, can, do, down, even, for, from, had, has, have, her, his,*
(25-36) *if, in, into, is, it, its, may, more, must, no, not, now,*
(37-48) *of, on, one, only, or, our, so, some, such, than, that, the,*
(49-60) *their, then, there, things, this, to, up, was, were, what, when, which,*
(61-63) *who, with, would*

Our goal is to construct for each justice a variability measure of opinions. The intuition behind this measure, as stated in the introduction, is that the greater the variability in writing style, the more likely that justice has delegated at least part of her writing responsibilities to her law clerk.

Because we are counting the appearance of function words, we adopt a chi-squared approach. The chi-square statistic allows us to test the distribution of the observed count of words against a theoretical or expected distribution. In our case, we will compare the observed count with the null hypothesis that the variability follows a chi-squared distribution, i.e., the total count of each function word is equally likely to occur in any of the total number of opinions for each justice.

The following paragraphs more formally describe our approach. The 63 function words are numbered from $j = 1$ to $j = 63$. Suppose a given justice has written opinions numbered from $i = 1$ to $i = K$. Let w_i be the total number of words in judgment i and c_{ij} be the number of times that function word j appears in judgment i. We further define $e_{ij} = w_i\left(\frac{c_{1j} + c_{2j} + \ldots + c_{Kj}}{w_1 + w_2 + \ldots + w_K}\right)$, where e_{ij} is the expected number of times that function word j would have appeared in judgment i. The null hypothesis is that the total number of $c_{1j} + c_{2j} + \ldots + \underline{c}_{Kj}$ appearances of j were each equally likely to occur in any of the total number of $w_1 + w_2 + \ldots + \underline{w}_K$ words in all of justice K's combined judgments.

The chi-squared statistic is the following:

$$chisq = \sum_{i=1}^{K} \sum_{j=0}^{63} \frac{(c_{ij} - e_{ij})}{e_{ij}}$$

Under the null hypothesis, *chisq* should follow a chi-squared distribution with (63+1-1)(K-1) = 63(K-1) degrees of freedom, and therefore mean 63(K-1). This summation also includes $c_{io} = w_i - c_{i1} - \ldots - c_{iK}$, the number of words in judgment i which are *not* function words.[31]

Our variability measure is accordingly

$$V\ score = \frac{chisq}{df} = \frac{chisq}{63(K-1)},$$

[31] The "+1" arises from the c_{io} terms.

which should approximate 1 under the null hypothesis, and exceed 1 for opinions that collectively exhibit greater variability in writing style.

There are, of course, other ways of constructing variability scores, many of which we also tried but ultimately decided against using. Because opinions vary in length, using raw counts c_{ij} by themselves would not be informative. One alternative would be to evaluate the fraction of words in judgment i which use function word j, i.e., $f_{ij} = c_{ij}/w_{ij}$. Under this approach, a given justice has a fixed unknown propensity p_j for using j, independently for each word of each opinion. Accordingly, the distribution of c_{ij} of j in i is Binomial (w_i, p_j), so that f_{ij} has mean p_j and variance $p_j(1-p_j)/w_i$, which would depend on individual propensities p_j and w_i.[32] We decide against this analysis because it is dependent on w and p, meaning the variability score will be biased downward for justices who write shorter opinions. Similarly, while it is possible to modify the measure such that is independent of w and p,[33] it would still be based on an imperfect estimate if the propensities p_j.[34]

It is worth noting that the chi-squared values may be less stable, and therefore less meaningful when many of the expected cell counts for given words approximate zero. We control for this in part by excluding function words that have a very low frequency in majority opinions, as well as opinions shorter than 250 words. A small fraction of the cell counts – both expected and observed – have a count less than one. It is possible to correct for this, such as a Yates correction, but is typically used for 2 x 2 tables. As a check, we recomputed our V score omitting all cells with a very small expected cell count: the scores reduced slightly in a consistent format, but the bootstrap tests of significance, *infra*, were unchanged. Accordingly, we leave unchanged our definition of V score.

---

[32] The variability measure would then be the sum of sample standard deviations, i.e., $V = \sum_{j=1}^{63} sd(f_{1j}, f_{2j}, f_{3j}, ...f_{Kj})$, where larger standard deviations reflect a more variable writing style.

[33] Because f_{ij} has mean 0 and variance $p_j \frac{(1-p_j)}{w_i}$, this is also true for $f_{ij} - \mu_j$, where $\mu_j = \frac{c_{1j} + c_{2j} + \ldots + c_{Kj}}{w_1 + w_2 + \ldots + w_K}$ is our best estimate of p_j. The quantity $r_{ij} = w_i^{1/2}(f_{ij} - \mu_j)$ has mean 0 and variance $p_j(1-p_j)$, resulting in $V = \sum_{j=1}^{63} sd(r_{1j}, r_{2j}, r_{3j}, ...r_{Kj})$.

[34] We can eliminate formal dependence on both w and p by creating $V = \sum_{j=1}^{63} sd(q_{1j}, q_{2j}, q_{3j}, ..q_{Kj})$, where $r_{ij} = \frac{w_i^{1/2}(f_{ij} - \mu_j)}{(\mu_j(1-\mu_j))^{1/2}}$. The term μ generates uncertainty where p is close to zero (in the most extreme case, where $\mu_j = 0$, r_{ij} is undefined.

For these reasons, we ultimately chose the chi-squared approach, which serves as the foundation for our analysis.[35] The first stage determines whether the justices' variability scores are distinguishable from our null hypothesis. Establishing that they are, we then use a bootstrap approach to determine whether their variability scores are statistically distinguishable from one another. Third, we then construct a linear classifier to test whether our model is able to accurately predict authorship, allowing us to construct variability scores for each justice, over their entire tenure as well as particular periods. It also allows us to directly test variability in pairwise comparisons of justices using bootstrap simulations. Finally, we test that accuracy of our model in predicting authorship through *leave-one-out cross-validation*. We describe these latter analytic approaches in detail in Part IV.

Part III: Data

Our data consists of the written opinions from the United States Supreme Court, which we obtained from *justia.com*, a website that provides free access to federal government documents. Specifically, *justia* includes all published Supreme Court opinions, from 1789 through the current term.

In order to statistically analyze the opinions, we wrote software in C and Unix[36] that downloaded the decisions directly from *Justia*. Among other things, our program converts the html pages into plain text and culls from the opinion(s) any text not written by a justice, such as headnotes, synopses, and other notes. In the process, the program separates majority from concurring and dissenting opinions, a surprisingly challenging task attributable to changing conventions of how the Court demarcate various types of opinions.

For both methodological and substantive reasons, we focus on majority opinions. Although more common in recent years, dissents were relatively historically rare until 1943.[37] Concurrences, then as well as now, are less common than dissents.[38] Compared with majority opinions, dissents and concurrences are also typically shorter, often significantly so, creating instability in the textual analysis. In addition, we exclude any unsigned opinions, which typically arose in the form of *per curiam* opinions, court orders, decrees, or motions. We also exclude any opinions shorter than 250 words.

---

[35] For a technical discussion of the statistical approach, *see* Jeff S. Rosenthal and Albert H. Yoon, *Detecting Multiple Authorship of United States Supreme Court Legal Decisions Using Function Words*, forthcoming __ ANN. APP. STAT. __ (2011).

[36] On *Justia*, for most volumes, the majority opinion and any dissenting or concurring opinions for each case are contained within a single html file. Our software program identifies the beginning and end of each majority opinion. We have made the software for downloading and analyzing these texts available. For a full description of the software program, see http://probability.ca/usscj/README.

[37] Prior to 1943, approximately 10 percent of decisions included a dissent. Subsequently, in most Court terms, over 50 percent of decisions include a dissent. *See* LEE EPSTEIN, ET AL, THE SUPREME COURT COMPENDIUM, Table 3.2, 211-15 (2003).

[38] *See id.* at Table 3.3, 216-220.

We analyzed opinions spanning the entire history of the Court, which includes 115 Justices. Given this sizable list, for our analysis of individual justices we report the most recent forty-one justices, based on their year of departure from the Court.

Part IV: Results

In this section we describe the variability scores (hereafter V scores) for each justice, which allow us to determine the distinctiveness of justices' writing style generally, and against one another. We first establish whether these scores are statistically distinguishable from the null hypothesis, then from one another. Afterwards, we test our model's ability to accurately predict authorship in pairwise comparisons of justices.

Variability (V) Scores: Table 2 lists the 41 most recent justices, those who joined the Court since 1925, providing brief demographic information and their V scores. The V scores represent each justice's overall variability scores based on all of his or her majority opinions.[39] For reasons described earlier in Part II, we exclude dissenting and concurring opinions, as well as any majority opinion of length less than 250 words.

The interpretation of the V score is as follows: the higher a justice's V score, the more variable her writing style, based on the use of function words in Table 1. A higher V score is consistent with the view that the justice relies more on her law clerks in writing the opinions. It bears repeating that our analysis cannot prove this result, and is open to competing explanations. For example, even in a world where every justice writes her own opinions, one would expect that some justices simply have more variable writing styles than others. This heterogeneity is also likely true in a world where every justice relies on law clerks to assist in writing opinions. The variability in V scores may also reflect differences in law clerks, namely their ability to mimic their justices' writing style.[40]

What is the validity of the V scores? The first is to test our construction of the null hypothesis that that the justices writing style follows a uniform and random distribution of function words. To test this, we randomly generate 200 pseudo-documents each consisting of 2000 independently and randomly generated words. Each word was chosen to be a non-function word with probability 70 percent, and uniformly selected from the list of function words in Table 1 with probability 30 percent. The null is a V score that approximates 1. Repeating this experiment 10 times, we produced a mean V score of 1.004622 with a standard deviation of 0.001702, consistent with our null having a true mean equal to 1.

[39] We provide a list of justices departing prior to 1941 in Appendix Table 2A.

[40] Robert O'Neill, one of Justice Brennan's clerks early in his tenure, commented that he and Richard Posner – his co-clerk – were allowed to draft opinions throughout their clerkship because of their ability to mimic Brennan's style or voice. *See* PEPPERS, *supra* note __, at 158.

Table 2
Variability Score
Current and Departed Justices Since 1941

Number	Justice	Appointing President	Year Joined Court	Year Left Court	Age at Retirement	Years on Court	Opinions	Average Word Length	V Score
1	Sonia Sotomayor	Obama	2009			1	5	4259	3.00
2	Samuel A. Alito	Bush, G. W.	2006			4	30	4274	3.32
3	John G. Roberts	Bush, G. W.	2005			5	32	4471	3.06
4	Stephen G. Breyer	Clinton	1994			16	122	3775	3.04
5	Ruth Bader Ginsburg	Clinton	1993			17	137	4489	3.57
6	Clarence Thomas	Bush, G. H. W.	1991			19	151	4041	3.57
7	Anthony M. Kennedy	Reagan	1988			22	197	5083	3.75
8	Antonin Scalia	Reagan	1986			24	228	4371	3.08
9	John Paul Stevens	Ford	1975	2010	90	35	378	4397	3.39
10	David H. Souter	Bush, G. H. W.	1990	2009	70	19	155	5370	3.70
11	Sandra Day O'Connor	Reagan	1981	2006	76	25	296	5100	3.92
12	William H. Rehnquist	Reagan	1972	2005	81	33	451	3672	3.12
13	Harry A. Blackmun	Nixon	1970	1994	86	24	313	4403	3.70
14	Byron Raymond White	Kennedy	1962	1993	76	31	469	4826	3.71
15	Thurgood Marshall	Johnson, L.	1967	1991	83	24	311	3634	3.38
16	William J. Brennan	Eisenhower	1956	1990	84	34	451	3762	3.34
17	Lewis F. Powell	Nixon	1972	1987	80	15	252	3818	3.27
18	Warren Earl Burger	Nixon	1969	1986	79	17	256	3742	3.13
19	Potter Stewart	Eisenhower	1958	1981	66	23	300	3265	3.28
20	William Orville Douglas	Roosevelt, F.	1939	1975	77	36	510	2140	2.54
21	John Marshall Harlan	Eisenhower	1955	1971	72	16	164	3533	2.92
22	Hugo Lafayette Black	Roosevelt, F.	1937	1971	85	34	470	1914	2.46
23	Abe Fortas	Johnson, L.	1965	1969	59	4	39	3261	2.62
24	Earl Warren	Eisenhower	1953	1969	78	16	165	3224	3.16
25	Tom Campbell Clark	Truman	1949	1967	68	18	215	2606	2.74
26	Arthur Joseph Goldberg	Kennedy	1962	1965	57	3	36	3714	3.15
27	Charles Evans Whittaker	Eisenhower	1957	1962	61	5	42	2597	3.21
28	Felix Frankfurter	Roosevelt, F.	1939	1962	80	23	250	2880	2.69
29	Harold Hitz Burton	Truman	1945	1958	70	13	96	3022	3.69
30	Stanley Forman Reed	Roosevelt, F.	1938	1957	73	19	232	3186	2.77
31	Sherman Minton	Truman	1949	1956	66	7	64	1580	2.27
32	Robert Houghwout Jackson	Roosevelt, F.	1941	1954	62	13	147	2881	2.52
33	Fred Moore Vinson	Truman	1946	1953	63	7	76	3305	2.95
34	Wiley Blount Rutledge	Roosevelt, F.	1943	1949	55	6	65	4484	3.41
35	Frank Murphy	Roosevelt, F.	1940	1949	59	9	129	2189	2.73
36	Harlan Fiske Stone	Roosevelt, F.	1925	1946	74	21	444	2511	2.96
37	Owen Josephus Roberts	Hoover	1930	1945	70	15	288	2259	2.68
38	James Francis Byrnes	Roosevelt, F.	1941	1942	63	1	16	2387	2.55
39	Charles Evans Hughes	Hoover	1930	1941	79	11	390	2811	3.06
40	Harlan Fiske Stone	Coolidge	1925	1941	69	16	444	2511	2.96
41	James Clark McReynolds	Wilson	1914	1941	79	27	464	1245	2.08

Note:Justices ordered in reverse chronological order, first by year of departure, then by year of appointment. Justices who departed the Court prior to 1941 are listed in the Appendix, Table A1.

The V scores range from a low of 2.08 from Justice McReynolds to a high of 3.92 for Justice O'Connor. For each justice, the V score is much larger than it would be under the null hypothesis that the function words are truly distributed uniformly and randomly. For example,

Justice Breyer has a V score of 3.04, which reflects a *chisq* statistic of 3.04 x 63 x (121-1) = 23,003.58. The null hypothesis has a *chi-squared* distribution of 63 x 120 = 7,560. Justice Breyer's V score corresponds to a p-value less than 0.0000001, which allows us to reject the null hypothesis, based on the function words, that Breyer's writing follows a uniform and random distribution. We can reject the null hypothesis for all the justices in Table 2, as well as all justices throughout the history of the Court for which we could produce a V score.[41]

The results also support anecdotal accounts of justices' approaches to opinion writing. For example, Justice Douglas has a V Score of 2.54, among the lowest of the justices. This score is consistent with scholarly accounts[42] as well as Justice Douglas's own claim that he wrote his own opinions.[43] Similarly, Justice Scalia has a reputation for writing his own opinions;[44] his V score of 3.08, while higher than Douglas, is among the lowest among current justices. Justice Holmes, listed in the appendix, had the lowest V score of all justices: 1.76. This comports with his reputation as one who used his clerks sparingly and primarily for minor tasks such as cite checking.[45] Similarly, Justice Cardozo (appendix), particularly renowned for his writing, had a V score of 2.30.

Several justices with high V scores similarly support anecdotal and historical accounts. Justice Blackmun's V score (3.70) supports views by some historians that Blackmun delegated much of the opinion writing to his clerks, more so than his contemporaries.[46] Interestingly, the justices with the highest V scores were Justices O'Connor (3.92) and Kennedy (3.75). Journalists and scholars have written at length about these justices' influence on the Court as being pivotal, or swing, voters,[47] but have largely ignored the relationships of these justices to their clerks. Whether their higher V scores are attributably to their ideological position at the median or their personal approach to opinion writing is beyond the scope of this article. It is

---

[41] Because several justices on the Court prior to 1800 did not write any opinions in which they were attributed authorship, we could not calculate a V score for them.

[42] *See* WARD AND WEIDEN, *supra* note __, at 205.

[43] Transcriptions of Conversations between Justice William O. Douglas and Professor Walter F. Murphy, Cassette No. 3: December 20, 1961, Princeton University Library, 1981 (stating "I have written all my own opinions.")

[44] *See* EDWARD LAZARUS, CLOSED CHAMBERS 271 (1998) (distinguishing Justice Scalia – and Jusitce Stevens – among the other justices in the opinion-writing process).

[45] *See* WARD AND WEIDEN, *supra* note __, at 35.

[46] *See* Garrow, *supra* note __.

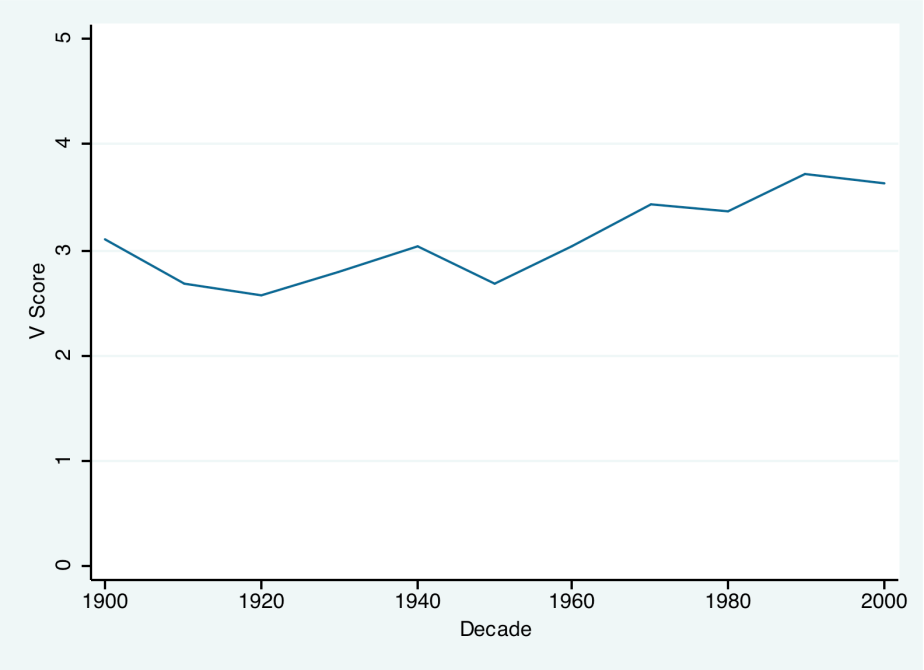
[47] *See, e.g.,* Douglas M. Parker, *Justice Kennedy: The Swing Voter and His Critics*, 11 Greenbag 317 (2008) (describing Justice Kennedy as the successor to Justice O'Connor as the swing voter on the Court); LAZARUS, *supra* note __, at 209 (describing Justice O'Connor's influence as the "crucial swing" vote during the 1988-89 term). Lynn A. Baker, *Interdisciplinary Due Diligence: The Case for Common Sense in the Search for the Swing Justice*, 70 S. CAL. L. REV. 187, 207 (1996) (describing Justices Kennedy and O'Connor's influence as swing voters).

worth noting, however, that Justice Byron White, another noted swing justice, had the third highest V score (3.71) of all justices.[48]

Some V scores did surprise us. Justice Thurgood Marshall's score of 3.38 were comparable to his contemporaries on the Court, most notably Justice Brennan (3.34). His score runs counter to accounts that Justice Marshall largely delegated opinion writing to his clerks. Justice Ginsburg's score of 3.57 appears unremarkable, falling roughly in the middle of her contemporaries. But it is markedly higher than Justices Breyer (3.04) or Scalia (3.08), suggesting that she relies more on her clerks in writing opinions than do her colleagues who were formerly law professors.

We can aggregate the justices' individual V scores to show trends of the Court over time, as reported in Figure 1 and Table 3.

Figure 1
Variability Score
By Decade (1900s-2000s)



[48] B. Drummond Ayres, Jr., "The 'Swing' Justice," N.Y. TIMES, June 30, 1972.

Table 3
Variability Score
By Chief Justice (1910-2009)

Chief Justice	Start Yr	End Yr	Start Vol	End Vol	Total Years	Opinions	Average Word Length	V Score
Roberts	2005	2009	546	559	4	284	4288	3.59
Rehnquist	1986	2005	478	545	19	1803	4503	3.61
Burger	1969	1986	395	478	17	2229	4067	3.47
Warren	1953	1969	346	395	16	1497	2770	2.91
Vinson	1946	1953	329	346	7	742	2726	3.02
Stone	1940	1945	314	328	5	711	2705	2.96
Hughes	1930	1941	280	313	11	1811	2178	2.79
Taft	1921	1930	257	280	9	1593	1815	2.58
White	1910	1921	218	256	11	2379	2020	2.64

A couple of patterns emerge. The V scores are lower and generally stable for the period 1900-1950, and steadily increase beginning in 1950.In the first half of the 20th century, law clerks served a largely administrative role.[49] This comports with historical accounts that beginning in the 1950s, clerks began to take a more active role in substantive matters of the Court, including drafting of opinions.

While informative, a justice's aggregate V score provides an incomplete account of her writing style. An aggregate score may reflect consistency over time, or it may mask an upward or downward trend or even considerable changes from one year to the next. We examine the justices' V scores on a year-by-year basis, which allow us to observe time trends, and to calculate variation over time.

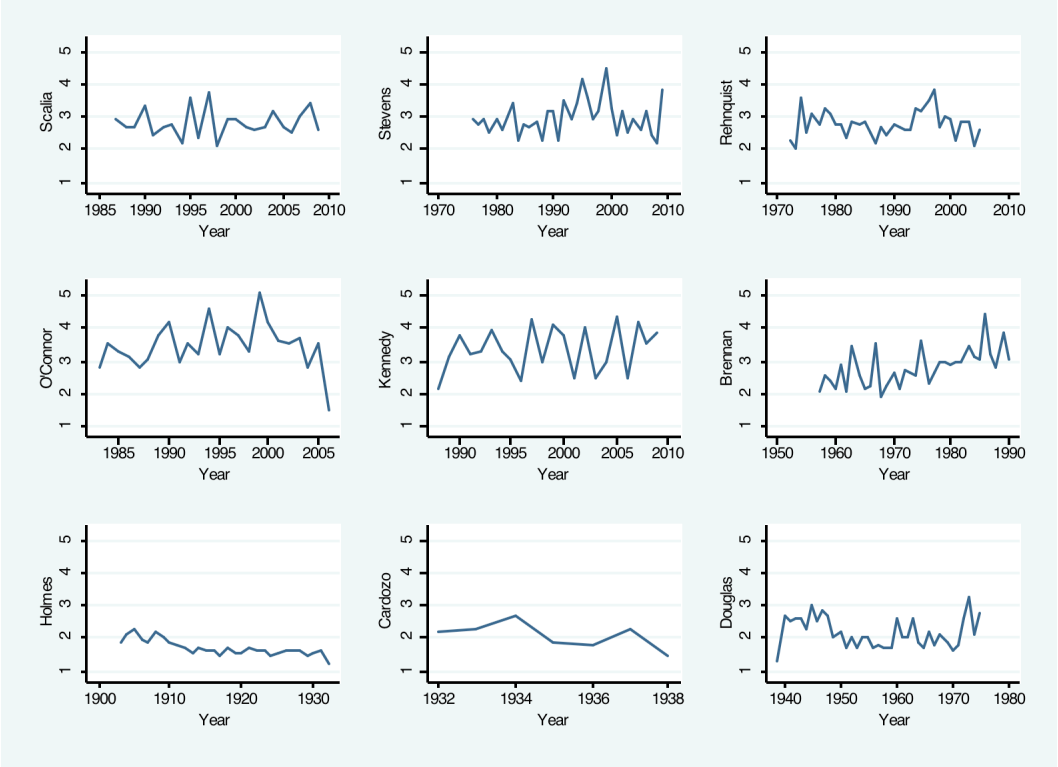
We report a sample of justices in Figure 2. Justice Holmes, with an aggregate V score of 1.76, showed remarkable consistency from year to year, with a standard deviation of only 0.22. Justices Cardozo, Douglas, Rehnquist and Scalia also had relatively low V scores, approximately 0.40. By contrast, O'Connor and Kennedy showed greater variability in V scores across years, each with standard deviations of approximately 0.68. Stevens and Brennan, with standard deviations of 0.53 and 0.57, respectively, fall between these groups. Their V scores exhibit a different pattern. Stevens' scores have three distinct periods: in his first dozen years (1976-1988), Stevens had an average V score of 2.76, with a standard deviation of 0.31; from 1989 through 2000, he had an average V score of 3.35 and a standard deviation of 0.60. From 2001 onward, Stevens' V scores resembled his early years on the Court: averaging 2.82 with a standard deviation of 0.52. Justice Brennan, by contrast shows an upward trend in V scores by decade, with fairly constant standard deviations (approximately 0.50) during each period.

Figure 2 provides closer insight into the writing relationships that justices have with their law clerks. Because law clerks typically serve for only one year, the annual V score arguably provides a provides a measure of a justice's reliance on them. Assuming that clerks vary in writing style from one another, it logically follows that justices who rely more on law clerks would experience greater changes from year to year than those who are known to do their own

[49] *See* PEPPERS, *supra* note / /, at 83-145, describing law clerks during this period as legal assistants.

writing. Holmes, Cardozo, and Douglas were all known to do their own writing:[50] the year by year V scores support this belief. In contrast, by this measure it appears that O'Connor and Kennedy rely more on their clerks. Their year-to-year scores have the highest standard deviation among the justices included in Figure 2.

Figure 2
Year-to-Year V Scores
Select Justices



We recognize that the validity of V scores, as a measure of justices' reliance on their law clerks, is ultimately unverifiable. Justices' relationship with their clerks remains shrouded in secrecy, so we cannot definitively separate the justices' own writing styles from those of their clerks. One recent study has attempted to compare draft majority opinions with the final published versions, finding differences across justices.[51] Given that only a few justices' papers are in the public domain, we cannot evaluate this more systematically. More importantly, comparing versions of opinions leaves unanswered the authorship process generally.

We check the validity of the V scores by comparing them with two jurists known to write their own opinions: Richard A. Posner and Frank Easterbrook, both of the U.S. Court of Appeals for the Seventh Circuit. Judge Posner has described his writing process as one where he writes

[50] *See* Ward & Weiden, *supra* note __, at 58, 96, 114 (describing the clerkship responsibilities for each justice).

[51] See Wahlbeck, *supra* note __.

his own draft opinions, and asks his clerks to "make criticisms and do research and tie up the loose ends that I discovered in writing."[52] He then incorporates their feedback in his revision.[53] Easterbrook has similar relationship with his clerks, differing in one respect: he writes all of his opinions but allows each of his clerks to draft one opinion during the term.[54] Together, these judges provide a reasonable gauge to determine whether the V score provides a reasonable measure of writing variability.

Given that Posner and Easterbrook write their own opinions, one would expect their V scores to be relatively low and stable from one year to the next, on the theory that their writing style is not highly dependent on their law clerks. The summary statistics, reported in Table 4, support this view. Both Posner and Easterbook have V scores that are lower than any of their contemporaries on the Court and lower than all but a handful of the justices in Table 2. It is worth noting that while their average word length is shorter than most justices, they have both written many more opinions than any justice, lending support to our belief that the V scores provide a valid measure of writing variability.

Table 4
Variability Score
Judges Posner and Easterbrook

Years	Judge	Opinions	Average Word Length	V Score
1981-present	Richard A. Posner	2372	2714	2.60
1984-present	Frank Easterbrook	1651	2400	2.42

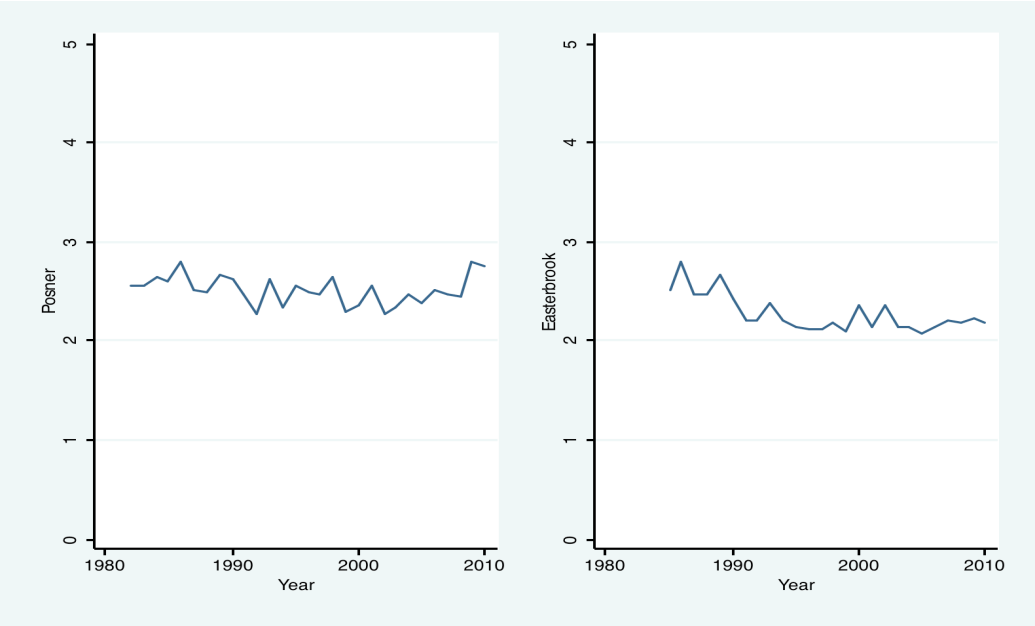
Figure 3 shows that the V scores for both Judge Posner and Judge Easterbrook V has remained stable during their tenure on the bench, particularly when contrasted with current and recent justices in Figure 2. For Judge Posner, the standard deviation in his V scores across years was 0.14; for Easterbrook, the standard deviation was 0.18. Their standard deviations scores were lower than any of the contemporary justices, which range from 0.43 (Scalia) to 0.69 (O'Connor). Judges Posner's and Easterbrook's V scores empirically support the judges' reputations for writing their own opinions, and suggest, by contrast, that their contemporaries on the Court rely more on their clerks.

[52] *See* Richard A. Posner, Diary, Slate, Jan. 15, 2002 (available at http://www.slate.com/id/2060621/entry/2060742/).

[53] *See id.*

[54] *See* Stephen J. Choi and G. Mitu Gulati, *Which Judges Write Their Opinions (And Should We Care)?*, 32 FLA. ST. U. L. REV. 1077 (2005). The authors are in the early stages with Judge Easterbrook on a project to discern whether it is possible to identify which opinions he allowed his law clerks to draft.

Figure 3
Year-to-Year V Scores
Judges Posner and Easterbrook



In addition, we examined a sample of the justices' dissents. Dissents are typically shorter than majority opinions, although still lengthy (averaging over 3000 words). While some justices write relatively few dissents (e.g., swing justices), other justices often write them more often than majority opinions (e.g., ideological justices of the minority coalition). Looking just at Stevens, Kennedy, and Scalia, we found that the V scores for dissents, when compared with their majority opinions, were systematically smaller. This contrast suggests that justices engage in less delegation in dissenting opinions than for majority opinions, although an intuitive explanation for this escapes us.

Tests of Significance: While the justices' V scores allow us to convincingly reject the null hypothesis that their writing style follows uniform and random distribution of function words, it does not establish whether these scores are meaningfully different from one another. There is no straightforward analytic test: because we reject the null hypothesis, the justices' V scores by definition do not follow a chi-squared distribution. Nor can we analytically determine what type of distribution it is.

It is possible, however, to determine the distribution through a *bootstrap* test. Bootstrap is a procedure of repeated sampling with replacement from a given sample. The intuition behind bootstrapping is that it replaces the unknown distribution with the empirically determined distribution.

With our data, we select, for each justice, 100 cases of authored majority opinions uniformly at random, with repetition.[55] For each sample of 100 cases, we compute the V score in

[55] Sampling could occur *without* replacement, but the relative low number of opinions by some justices would cause them to drop from our analysis under this approach.

the same manner. We then repeat this process 1000 times for each justice: this generates 1000 different possible V scores, depending on which 100 cases are drawn.

We can then use these values to do pairwise comparisons of justices. This process creates 1,000,000 (1000 x 1000) pairs of V scores. We then simply count the fraction of pairs in which the V score for Justice A is greater than for Justice B, which gives us an estimate of the *probability* that the V Score for Justice A is greater than that for Justice B, for a random selection of judgments. These pairings allow us to estimate the distribution function for the difference of the V score for Justice A minus Justice B, which we can then use to compute a 95% confidence interval for this difference. Accordingly, a confidence interval that is entirely positive or negative indicates that the differences in V scores between Justice A and Justice B are statistically meaningful and not due to chance.

To illustrate, Table 5 provides two examples of bootstrap comparisons. For each, the bootstrap V scores for each of the justices closely approximate their V scores in Table 3. In the first comparison, Justice O'Connor has a larger bootstrap V score (3.87) than Chief Justice Rehnquist (3.09): the probability that her V score produces a lower score than his is less than 0.0002, a statistically significant difference. In the second example, Justice Stevens (3.33) produces a lower V score than Justice Thomas (3.50), which occurs 0.7692 of the time. This difference, however, is not statistically significant.

Table 5
Sample Pairwise Bootstrap Comparisons

Bootstrap V score (O'Connor)	Bootstrap V score (Rehnquist)	P(O'Connor < Rehnquist)	95% C.I.: (Rehnquist-Ginsburg)
3.87	3.09	0.0002	(-1.242248, -0.347346)

Bootstrap V score (Stevens)	Bootstrap V score (Thomas)	P(Stevens < Thomas)	95% C.I.: (Stevens-Thomas)
3.33	3.50	0.7692	(-0.321370, 0.660145)

Table 6 produces a pairwise comparison of all justices during the last Rehnquist natural court, a period covering 1994 through 2005. The column labeled V Score provides the V score from Table 2, which in each instance closely approximates the bootstrap V scores.[56] For the remainder of the grid, the cell numbers report the probability that the row justice has a lower V score than the column justice. For example, the probability of Ginsberg having a lower V score than Justice Breyer is 0.0130. Across the diagonal, the V scores are mirror images. Accordingly, the probability of Justice Breyer having a lower V score than Justice Ginsburg is 0.9870. Any probability greater than 0.95 or less than 0.05 reflects a statistically significant difference in the bootstrap V scores.

[56] We choose to report the original V score because each bootstrap produces a slightly different V score, but all of which are close to the V scores reported in Table 3.

Table 6
V Score Bootstrap – Pairwise Comparison
Justices on Last Rehnquist Natural Court (1994-2005)

			Column Justice								
		V Score	Stephen G. Breyer	Ruth Bader Ginsburg	Clarence Thomas	Anthony M. Kennedy	Antonin Scalia	John Paul Stevens	David H. Souter	Sandra Day O'Connor	William H. Rehnquist
Row Justice	Stephen G. Breyer	3.04		0.9870	0.9881	0.9993	0.7753	0.9287	0.9965	0.9998	0.6973
	Ruth Bader Ginsburg	3.57	0.0130		0.5033	0.8066	0.0119	0.2303	0.6769	0.9148	0.0300
	Clarence Thomas	3.57	0.0119	0.4967		0.7996	0.0202	0.2308	0.6898	0.9263	0.0286
	Anthony M. Kennedy	3.75	0.0007	0.1934	0.2004		0.0004	0.0585	0.4013	0.7426	0.0013
	Antonin Scalia	3.08	0.2247	0.9881	0.9798	0.9996		0.8989	0.9946	0.9999	0.5934
	John Paul Stevens	3.39	0.0713	0.7697	0.7692	0.9415	0.1011		0.8771	0.9831	0.1390
	David H. Souter	3.70	0.0035	0.3231	0.3102	0.5987	0.0054	0.1229		0.7829	0.0088
	Sandra Day O'Connor	3.92	0.0002	0.0853	0.0737	0.2574	0.0001	0.0169	0.2171		0.0002
	William H. Rehnquist	3.12	0.3027	0.9700	0.9714	0.9987	0.4066	0.8610	0.9912	0.9998	

Note: V Score column represents V Score for each Justice. Remaining cells represent the P((Row Justice)<(Column Justice)).

The grid reveals that in a large fraction of pairings – 44 percent – the justices are statistically distinguishable from one another. The larger the difference in V scores, the more likely the bootstrap produces a statistically distinguishable score. Some justices, such as Justice Ginsburg and Justice Thomas, have V scores that are not statistically distinguishable from each other, based on our selection of function words. Because their V scores are identical, it is not surprising that the probabilities are very close to 0.50. Conversely, Justice Breyer, with the lowest V score of this cohort, produces bootstrap estimates that are statistically distinguishable from every other justice during this period, save Justice Scalia and Chief Justice Rehnquist. Collectively, 16 out of 36 of all the pairwise bootstrap comparisons were statistically significant. As one might expect, justices with V scores near the median of the Court were less distinguishable than those justices with relatively high or low V scores. Even when the scores were not statistically significant, in 33 of the 36 pairings the probability was either greater than 0.70 or less than 0.30.

Table 7
Within-Justice Comparison – Pre and Post-65
By Justice

	Before Age 65			After Age 65			Bootstrapping		
Justice	Opinions	Avg Word Length	V4	Opinions	Avg Word Length	V4	P(pre65< Post65)	lower 95 CI	upper 95 CI
Sonia Sotomayor									
Samuel A. Alito									
John G. Roberts									
Stephen G. Breyer	70	3938	3.17	46	3456	2.81	0.0465	-0.75	0.05
Ruth Bader Ginsburg	41	4918	3.57	90	4373	3.52	0.5113	-0.42	0.45
Clarence Thomas	151	4041	3.57						
Anthony M. Kennedy	134	4827	3.64	57	5740	3.91	0.8432	-0.22	0.68
Antonin Scalia	152	4520	3.02	71	4051	3.05	0.5265	-0.31	0.33
John Paul Stevens	121	3939	3.10	244	4689	3.42	0.8793	-0.22	0.81
David H. Souter	114	5761	3.82	35	4283	3.10	0.0003	-1.29	-0.27
Sandra Day O'Connor	202	5063	3.73	88	5208	4.28	0.9661	-0.04	1.04
William H. Rehnquist	277	3608	2.95	164	3827	3.16	0.8766	-0.14	0.54
Harry A. Blackmun	21	4672	4.14	280	4341	3.51	0.0719	-1.06	0.17
Byron Raymond White	267	4960	3.74	187	4443	3.44	0.1335	-0.84	0.22
Thurgood Marshall	56	2948	3.16	246	3773	3.31	0.7430	-0.30	0.61
William J. Brennan	168	3099	2.81	276	4206	3.48	0.9988	0.24	1.15
Lewis F. Powell				240	3832	3.27			
Warren Earl Burger	24	5139	3.16	220	3570	3.08	0.5720	-0.40	0.50
Potter Stewart	272	3227	3.25	16	3722	3.11	0.1262	-0.83	0.16
William Orville Douglas	367	2215	2.54	136	1959	2.42	0.2433	-0.56	0.26
John Marshall Harlan	31	2860	2.53	68	4075	3.09	0.9972	0.16	0.97
Hugo Lafayette Black	260	1860	2.48	206	1997	2.37	0.4681	-0.49	0.34
Abe Fortas	39	3261	2.62						
Earl Warren	25	1682	2.39	136	3550	3.27	1.0000	0.50	1.41
Tom Campbell Clark	104	2080	2.49	35	3447	3.12	0.9992	0.20	0.93
Arthur Joseph Goldberg	36	3714	3.15						
Charles Evans Whittaker	42	2597	3.21						
Felix Frankfurter	104	2458	2.50	141	3204	2.79	0.9325	-0.09	0.68
Harold Hitz Burton	51	3458	4.07	38	2642	3.36	0.0888	-1.45	0.24
Stanley Forman Reed	157	3216	2.75	62	3094	2.61	0.2389	-0.46	0.22
Sherman Minton	54	1611	2.30	9	1463	1.96	0.0000	-0.69	-0.32
Robert Houghwout Jackson	147	2881	2.52						
Fred Moore Vinson	76	3305	2.95						
Wiley Blount Rutledge	65	4484	3.41						
Frank Murphy	129	2189	2.73						
Harlan Fiske Stone	248	1977	2.63	183	3222	3.33	0.9983	0.23	1.19
Owen Josephus Roberts	195	2138	2.61	84	2567	2.84	0.8198	-0.22	0.62
James Francis Byrnes	16	2387	2.55						
Charles Evans Hughes	139	2800	2.89	251	2817	3.04	0.7124	-0.33	0.59
Harlan Fiske Stone	248	1977	2.63	183	3222	3.33	0.9988	0.23	1.17
James Clark McReynolds	261	1188	2.05	190	1302	2.10	0.6541	-0.29	0.36

Note: We do not analyze justices who have been on the Court fewer than 5 years.

In addition to comparing across justices, the bootstrap test also allows us to compare V scores *within* justices. This analysis allows us to examine how an individual justice may have changed over time. Political scientists have commented how justices undergo ideological shifts

during their tenure on the Court.[57] Although our variability measure – relying on common function words – is non-ideological, changes in ideology may coincide with changes in reliance on law clerks.[58]

While there are numerous ways to evaluate change over time, we first examine justices' writing before and after their 65th birthday. Sixty-five is an important benchmark for multiple reasons: it represents the typical retirement age among most workers and also represents the age at which many justices vest in their judicial pensions.[59] Moreover, some scholars have argued that justices too often remain on the Court long past their productive years, to the point of "mental decrepitude."[60] A dramatic change in V score before and after age 65 may suggest that older justices are delegating more of their writing to their clerks.

Table 7 reveals that a majority of justices have a higher V score *after* age 65 than before. Justice O'Connor, for example, had a V score of 3.83 pre-65, and 4.28 post-65. Justices Kennedy, Scalia, Stevens, Rehnquist, Marshall, Brennan, Harlan, Clark, Frankfurter, and Stone followed this trend. Other justices, such as Breyer, Ginsburg, Souter, Blackmun White, Burger, Stewart, Douglas, Black, Burton, Reed and Minton experienced a reverse trend. In most instances, however, the difference across periods was not statistically significant, suggesting that older age, at least by this measure, does not manifest itself in greater variability in writing style.

Because pre- and post-65 is a broad measure of change in the Court, it may make sense to look at more narrow time periods. For example, it might make sense to contrast justices' first five years with their final five years on the Court. The intuition behind examining these periods is that, even for justices with prior judicial experience, the Court signifies a big adjustment. The caseload and writing requirements differ from any other Court. A docket that is nearly entirely discretionary and dedicated to new issues of law likely creates greater intellectual challenges than many justices encountered in their earlier tenure as judges or in legal practice.

---

[57] *See, e.g.,* Lee Epstein, et al, *Ideological Drift Among Supreme Court Justices: Who, When, and How Important?*, 101 N.W. L. Rev. 1483 (2007).

[58] Chief Justice Rehnquist famously argued in 1957, shortly after clerking himself on the Supreme Court, how clerks, predominantly from the political left, influence how justices make decisions.William H. Rehnquist, "Who Writes Decisions of the Supreme Court," U.S. NEWS & WORLD REP., Dec. 13, 1957, at 74-75

[59] The actual age at which justices, and all Article III judges, vests in their pension, is determined by statute, 28 U.S.C. Sec. 371. The current requirements, referred to as the *Rule of 80*, requires that the justice be at least 65 years of age and served at least 10 years, and that the age at retirement and years of service, when combined, be at least 80. *See* Albert H. Yoon, *Pensions, Politics, and Judicial Tenure: An Empirical Study of Federal Judges, 1869-2002* 8 AMER. L. & ECON. REV. 143, 147 (2003).

[60] *See* David J. Garrow, *Mental Decrepitude on the U.S. Supreme Court: The Historical Case for a 28th Amendment*, 67 U. CHI. LAW. REV. 995 (2000).

Table 8
Within-Justice Comparison – First 5 and Last 5 Years
Current and Departed Justices Since 1925

		First 5 Years on Court			Last 5 Years on Court			Bootstrapping		
Number	Justice	Opinions	Avg Word Length	V4	Opinions	Avg Word Length	V4	P(First5< Last5)	lower 95 CI	upper 95 CI
1	Sonia Sotomayor									
2	Samuel A. Alito									
3	John G. Roberts									
4	Stephen G. Breyer	34	4131	2.89	33	3755	3.01	0.7634	-0.18	0.40
5	Ruth Bader Ginsburg	41	4918	3.57	32	3682	3.32	0.1165	-0.67	0.16
6	Clarence Thomas	39	4766	3.77	36	3310	3.59	0.1942	-0.73	0.28
7	Anthony M. Kennedy	59	4410	3.40	26	6921	3.98	0.9940	0.10	0.87
8	Antonin Scalia	55	4224	3.11	37	4226	3.09	0.3785	-0.37	0.25
9	John Paul Stevens	50	3705	3.03	27	4319	3.41	0.8928	-0.18	0.79
10	David H. Souter	39	5327	3.83	35	4283	3.10	0.0000	-1.06	-0.38
11	Sandra Day O'Connor	64	5122	3.57	35	4287	3.95	0.9001	-0.18	0.79
12	William H. Rehnquist	78	4025	2.97	40	3415	2.78	0.1057	-0.58	0.12
13	Harry A. Blackmun	49	4661	3.92	52	4649	3.29	0.0208	-1.16	-0.02
14	Byron Raymond White	51	2997	2.89	77	4813	3.29	0.9946	0.09	0.77
15	Thurgood Marshall	38	2890	3.11	71	3399	2.86	0.1490	-0.63	0.17
16	William J. Brennan	51	2779	2.45	74	4535	3.85	1.0000	1.00	1.75
17	Lewis F. Powell	78	4096	3.49	87	3301	2.99	0.0152	-0.91	-0.04
18	Warren Earl Burger	55	4628	3.29	82	3035	3.05	0.1793	-0.65	0.24
19	Potter Stewart	46	3000	2.82	70	3629	3.29	0.9911	0.09	0.91
20	William Orville Douglas	105	2535	2.69	57	2336	2.87	0.6820	-0.42	0.57
21	John Marshall Harlan	45	2725	2.55	50	4617	3.35	0.9998	0.33	1.21
22	Hugo Lafayette Black	73	1815	2.49	57	2385	2.70	0.8914	-0.12	0.52
23	Abe Fortas	39	3261	2.62	39	3261	2.62	0.5080	-0.34	0.35
24	Earl Warren	44	1927	2.30	48	3981	3.50	1.0000	0.65	1.67
25	Tom Campbell Clark	47	2098	2.42	60	3537	3.06	0.9958	0.15	1.14
26	Arthur Joseph Goldberg	36	3714	3.15	36	3714	3.15	0.4964	-0.34	0.35
27	Charles Evans Whittaker	42	2597	3.21	39	2718	3.26	0.6268	-0.28	0.39
28	Felix Frankfurter	68	2488	2.52	43	4537	3.22	0.9973	0.17	1.13
29	Harold Hitz Burton	28	4137	4.88	38	2642	3.36	0.0104	-2.03	-0.51
30	Stanley Forman Reed	75	2658	2.49	39	2984	2.50	0.5266	-0.30	0.33
31	Sherman Minton	45	1694	2.33	43	1475	2.09	0.0120	-0.43	-0.03
32	Robert Houghwout Jackson	59	2923	2.59	46	2805	2.24	0.0059	-0.59	-0.08
33	Fred Moore Vinson	51	3533	2.98	54	2824	2.84	0.2462	-0.53	0.26
34	Wiley Blount Rutledge	49	4453	3.37	52	4675	3.10	0.0709	-0.67	0.09
35	Frank Murphy	68	1897	2.35	67	2430	2.87	0.9987	0.18	0.83
36	Harlan Fiske Stone	108	1882	2.60	95	3555	3.37	0.9999	0.32	1.37
37	Owen Josephus Roberts	89	1913	2.57	84	2567	2.84	0.8770	-0.16	0.68
38	James Francis Byrnes	16	2387	2.55	16	2387	2.55	0.5133	-0.16	0.17
39	Charles Evans Hughes	259	2758	3.00	99	2702	2.98	0.4130	-0.50	0.39
40	Harlan Fiske Stone	108	1882	2.60	95	3555	3.37	0.9996	0.32	1.35
41	James Clark McReynolds	85	1296	2.02	57	1180	1.89	0.0834	-0.34	0.06

Note: Justices who have been on the court for fewer than five years are not analyzed. For current justices, the last five years on the Court are their most recent five years.

The results, reported in Table 8, reveal greater differences within the justices. Sixteen out of 28 justices had higher V scores during their first 5 years as compared with their last 5 years. In most of these instances, however, this change was not statistically significant. Some interesting results emerge: Justice Scalia, for example, has shown remarkable consistency: his V score was 3.11 during his first five years, and 3.09 during his most recent five years. Justice

Marshall, speculated to have relied more heavily on his law clerks during his first years on the Court,[61] actually had a slightly lower V score during his final years on the Court (2.86) compared with his first five (3.11). Other justices, by contrast, changed considerable across these periods: Justice Souter had a significantly higher V score (3.83) in his first five years than in his last five years (3.10); while his predecessor, Justice Brennan, had a markedly *lower* V score (2.45) in his first five years than in his last five years (3.85). These results suggest that Justice Souter relied less on his clerks during his tenure, and Justice Brennan more.

Authorship Identification: Thus far, our writing variability measure has established that the justices have writing styles that are statistically distinguishable from the null, and – through our bootstrap test - in many instances statistically more variable than other justices. We now turn our attention to answer the question of practical importance: whether it is possible to accurately predict authorship of judicial opinions.

The consensus from our informal inquiries with constitutional law scholars suggest that, while they might be able to identify authorship based on known passages, it was unlikely that they could discern authorship based on writing style alone. The task is all the more challenging when comparing justices of similar judicial ideology. In this section we test whether it is possible to use of function words to accurately identify authorship.

As with the bootstrap, we approach this question through a pairwise approach. We consider a particular pair of justices – e.g., Justice A and Justice B, and the universe of majority opinions in which the author is one of these justices. Consistent with hypothesis testing, to avoid overfitting of the data,[62] we partition the data into two separate parts: a training set and a testing set. We use the training set to develop a model for classifying judgments as being authored by either Justice A or Justice B. We reserve the testing data to determine the predictive ability of our model.

To test its accuracy, we use *leave-one-out cross-validation*: for each judgment written by either Justice A or B, the judgment is the test set, and all other judgments written by either Justice A or B to serve as the training set. We determine whether the model properly classifies the test judgment as belonging to Justice A or B. We repeat this process for each judgment, and count the number of accurate classifications over the number of judgments by Justice A and Justice B, respectively.

With a linear classifier, we let T be a training set consisting of all judgments by Justice A or B, with $|T| = n$, where $\boldsymbol{n}$ are the total number of opinions. We use the following linear regression model,

$$Y = x\beta + \varepsilon,$$

[61] Mark Tushnet, *Thurgood Marshall and the Brethren*, 80 Geo. L. J. 2109, 2112 (1992) (writing that "Marhsall relied more heavily on his law clerks for opinion writing than did the other Justices during the early years of his tenure").

[62] Overfitting of the data is where one constructs a model that fits well on the existing data but fails to effectively predict on new data.

in which ε is an *n* x 1 vector of independent errors with mean zero. Y represents an *n* x 1 vector of ±1: we assign a value of -1 for each judgment written by Justice A and a +1 for each judgment written by Justice B. The term x is an *n* x 64 matrix defined as:

$$\mathrm{x} = \begin{pmatrix} 1 & f_{1,1} & f_{1,2} & \cdots & f_{1,63} \\ 1 & f_{2,1} & f_{2,2} & \cdots & f_{2,63} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & f_{n,1} & f_{n,2} & \cdots & f_{n,63} \end{pmatrix},$$

where $f_{i,j}$ are the fraction of words in judgment *i* in the training set which are from function word *j*.

The least-squares estimate for β corresponds to the maximum likelihood estimate (MLE) if the errors (ε_i) are assumed to be independent and identically distributed (i.i.d.), defined as

$$\hat{\beta} = (x^T x)^{-1} x^T Y$$

where $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots \hat{\beta}_n)$. Given a test judgment with function words fractions of g_1, g_2, . . . g_{63}, we can calculate the linear fit value

$$\ell = \hat{\beta}_0 + \sum_{j=1}^{63} \hat{\beta}_j g_j$$

We classify Justice A as the author of the test judgment if $\ell < 0$, and otherwise assign it to Justice B. While there are other possible means of classification – e.g., neural networks or support vector machines – we choose the linear classifier for its parsimony and accuracy.[63]

Table 9 provides two examples of the linear classifier. When comparing Justices Breyer and Ginsburg, the linear classifier accurately predicted the author of Justice Breyer's opinions 93 percent of the time, and Justice Ginsburg's opinions 96 percent of the time. A comparison of Justices Thomas and Kennedy produces lower accuracy results, accurately predicting authorship

[63] We also tried a naïve Bayes classifier, which assumes that 1) the different fractions over function words *j* are conditionally independent; 2) conditioned on Justice A being the author of record, the condititional distribution of the fraction f_j of function word *j* appearing in the judgment is normal; and 3) the corresponding mean and variance are given by the sample mean and variance by Justice A in the training data. This classifier produces a log likelihood of Justice A having written the opinion, $loglike(A) = C - \sum_{j=1}^{63} \left(\frac{1}{2} \log(v_j) + \frac{(f_j - m_j)^2}{2v_j} \right)$ for some constant C, where m_j and v_j are the same mean and variance of the fraction of words which are reference word *j*, over all judgments in the training set written by Justice A. We computer the *loglike*(B) in the same manner. The naïve Bayes produces similar, but on average slightly less accurate, predictions to the linear classifier.

of Justice Thomas's opinions 78 percent of the time, and Justice Kennedy's opinions 88 percent of the time.

Table 9
Linear Classifier
Sample Comparisons

	Justice A	Justice B	Fraction Accurately Predict Justice A	Success Rate Predicting Justice A	Fraction Accurately Predict Justice B	Success Rate Predicting Justice B
1	Stephen G. Breyer	Ruth Bader Ginsburg	112/121	0.9256	130/136	0.9559
2	Clarence Thomas	Anthony M. Kennedy	118/151	0.7815	173/197	0.8782

Table 10 produces the results of the pairwise comparisons for the linear classifier for all justices from Chief Justice Rehnquist's last natural court, which spanned from 1993 through 2005. The lowest prediction rates involved opinions written by Thomas. It predicted only 60 percent when compared with Stevens, 51 percent with Rehnquist, and only 36 percent with O'Connor. By contrast, the model's highest prediction rates involved Breyer, achievining a minimum accuracy rate of 87 percent in each of the eight pairings.

We note two interesting points of interpretation from the linear classifier analysis. One, justices may have similar V scores yet have writing styles that are clearly distinguishable from one another. For example, Justices Ginsburg and Thomas have identical V scores of 3.57. The linear classifier, however, is able to predict with 83 percent accuracy authorship of Justice Ginsburg's opinions, and 82 percent accuracy authorship of Justice Thomas's opinions. The explanation lies in the construction of the V score: similar – or even identical – V scores can reflect large differences between justices in their frequency of various function words. For example, while Justice Ginsburg and Justice Thomas use the word *his* at comparable rates (0.001974 and 0.002176, respectively), Justice Ginsburg uses the word *her* (mean = 0.000736) nearly twice as often as Justice Thomas (mean = 0.000375).

Table 10
Linear Classifier
Justices on Last Rehnquist Natural Court (1993-2005)

Authoring Justice (Justice A)		V Score	Non-Authoring Justice (Justice B): Stephen G. Breyer	Ruth Bader Ginsburg	Clarence Thomas	Anthony M. Kennedy	Antonin Scalia	John Paul Stevens	David H. Souter	Sandra Day O'Connor	William H. Rehnquist
	Stephen G. Breyer	3.04		0.9256	0.8926	0.9587	0.9669	0.9256	0.9256	0.8760	0.9091
	Ruth Bader Ginsburg	3.57	0.9559		0.8309	0.9338	0.9559	0.8824	0.8382	0.7868	0.8309
	Clarence Thomas	3.57	0.8543	0.8212		0.7815	0.7748	0.6026	0.8212	0.3642	0.5099
	Anthony M. Kennedy	3.75	0.9695	0.9188	0.8782		0.9086	0.7665	0.9289	0.7208	0.7767
	Antonin Scalia	3.08	0.9868	0.9474	0.8246	0.8904		0.7763	0.8904	0.7807	0.7456
	John Paul Stevens	3.39	0.9788	0.9524	0.8836	0.9127	0.9074		0.9524	0.7989	0.7196
	David H. Souter	3.70	0.9613	0.8710	0.8387	0.8839	0.8774	0.8323		0.7677	0.7677
	Sandra Day O'Connor	3.92	0.9527	0.9358	0.8243	0.8750	0.8784	0.7534	0.9324		0.6959
	William H. Rehnquist	3.12	0.9889	0.9579	0.9224	0.9401	0.9069	0.7938	0.9534	0.7783	

Note: V Score column represents V score for all opinions. Remaining cells represent fraction that accurately predicts Justice A compared to Justice B.

Second, unlike in Table 6, the values across the diagonal are not necessarily symmetric. In other words, the fraction predicting Justice A's opinions versus Justice B need not be the same for the fraction predicting Justice B's opinions versus Justice A. For example, in a comparison of Justices Thomas and Stevens, the linear classifier predicts Justice Thomas as the author of his opinions with only 60 percent accuracy, but predicts Justice Stevens with 88 percent accuracy. This asymmetry is due to the shape of the probability distributions of the two justices and their degree of overlap. If one distribution largely overlaps another distribution, it is possible for the prediction to be much higher for one than the other.[64]

[64] To illustrate how symmetries arise, consider the following figure:

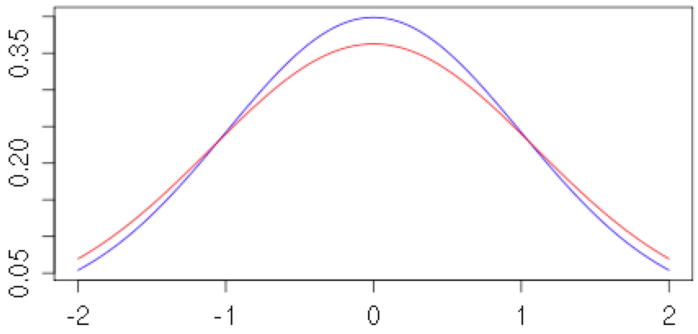
Overall, of the 72 possible pairings, the model achieved an accuracy rate of at least 70 percent in 69 pairings (representing over 95 of pairings). In 30 (42 percent) of the pairings, the accuracy rate exceeded 90 percent. Comparing these rates to a null hypothesis of authorship being randomly determined (i.e, 50 percent), our model appears to predict quite well. We achieved these results using function words, and it possible that a different algorithm – i.e, one that is more tailored towards words more common to legal writing – may produce even higher prediction rates.

The question of which justice authored a Supreme Court opinion is, admittedly, an academic exercise, since most majority opinions reveal the authoring justice. The purpose of this exercise is to show that the text of justices' opinions are statistically distinguishable from another, even in instances when their V scores are not statistically significantly more variable.[65] Our analysis shows that statistical analysis can meaningfully contribute to our understanding not merely how the Court votes,[66] but also how they write.

Part V: Discussion

This paper had two objectives. First, to show that it is possible to statistically evaluate justices' writings. Using only function words, we produced measures of justices' writing variability; in many instances these differences were. Moreover, our approach allows us to

---



The blue curve is a graph with mean 0 and variance one. The red curve is a graph with mean zero and variance 1.1. Most of the probability distribution falls between 1 and -1, where the blue curve is larger. In this example, approximately 70% of the points (opinions) chosen from either distribution will be classified as belonging to the blue curve. Accordingly, points from the blue distribution will be correctly classified about 70% of the time, while points from the red distribution will be correctly classified about 30% of the time.

[65] There may also be a practical application to the linear classifier when looking at *per curiam* opinions, for which the Court does not report the authoring justice. The linear classifier can offer be used to discern the likely author of these opinions, when analyzed in combination of known majority opinions.

[66] Most analysis of judicial politics examines the voting patterns of the justices based on the U.S. Supreme Court Judicial Database, created by Harold J. Spaeth.

accurately predict authorship, even when justices have similar or identical V scores. In over 95 percent of pairwise comparisons, our linear classifier model achieves an accuraty rate of at least 70 percent. In over 44 percent of comparisons, the classifier model exceeded 90 percent.

Our second, and more important, objective was to use textual analysis to better understand how justices produce opinions and their reliance on clerks. For the period 1900-1950, justices' V scores were low, consistent with historical accounts of clerks serving a primarily administrative role. Since 1950, however, their scores have steadily increased. This upward trend supports claims that justices are increasingly relying on their clerks in the opinion-writing process. In addition, many justices vary considerably in their V scores from one year to the next.

Moreover, our findings suggest that delegation of writing responsibilities to law clerks is highest among those considered swing justices. Justices Kennedy, O'Connor, Blackmun and White reported the highest V scores among all justices dating back to 1900. Whether being a swing justice fosters greater variability in writing, or whether the relationship is mere coincidence, the consequences are tangible and warrant closer examination. A large plurality of opinions each year are 5-4, often the most important cases of the term. Moreover, these cases are often assigned to swing justices.

We emphasize our earlier caveat that our analysis provides only circumstantial evidence of collaborative authorship. Low V scores may reflect that a justice does her own writing; conversely, the justice may delegate writing to her clerk but closely edits their writing, or that clerks effectively mimic their justices' writing style. High V scores may reflect greater delegation of writing responsibilities to the clerk, or merely that the justice does her own writing but simply has a higher baseline variability. While these alternative explanations are plausible, our separate analysis of opinions by judges Posner and Easterbrook, both known to write their own opinions, yields lower V scores, with lower variation from one year to the next. These results provide strong validity to our anlaysis.

Given our finding of greater delegation to clerks in the opinion-writing process, should we be concerned? Legal scholars have weighed in on the normative and positive implications of the judges relying on clerks in writing opinions. Richard Posner, not surprisingly, is critical of this trend:

> Most judges nowadays, because of heavy caseloads, delegate the writing of their judicial opinions to their clerks. It's a mistake to write on a number of grounds: The more you write, the faster you write; only the effort to articulate a decision exposes the weak joints in the analysis; and the judge-written opinion provides greater insight into the judge's values and reasoning process and so provides greater information—not least to the judge.[67]

Other scholars have argued that judges who write their own opinions are more influential than judges who rely heavily on clerks.[68] Mark Tushnet, a former law clerk for Justice Thurgood Marshall, has commented that "Surely, an opinion cannot carry the weight of a justice's prior

---

[67] *See* Richard A. Posner, Diary, Slate, Jan. 15, 2002 (available at http://www.slate.com/id/2060621/entry/2060742/).

[68] *See, e.g.,* William M. Landes, Lawrence Lessig, Michael E. Solimine, *Judicial Influence: A Citation Analyss of Federal Courts of Appeals Judges*, 27 J. LEGAL STUD. 271, 274 (1998).

public service when it is written by a recent law school graduate serving as the Justice's law clerks."[69]

Our analysis is silent on whether increased delegation to clerks has improved or harmed the quality of the Court's opinions. Answering this question presents its own challenges, on two fronts: one, the non-random process by which the Chief Justice (or Senior Associate Justice, if the Chief is not in the majority) assigns opinions; and two, no obvious metric for measuring the quality of the justices' writing. One point is clear: justices now write opinions that are longer and more variable along common function words. While much of this can be attributed to technological advances, it is not difficult to imagine that justices would write shorter, less variable opinions without the assistance from their clerks.

Our results also inform existing debates about the Court. For example, recent scholarship has criticized lifetime tenure for Supreme Court justices and proposed a constitutional amendment to impose term limits.[70] We do not find that older justices – i.e., those greater than 65 years old – have systematically larger V scores than in their younger years. Nor do we find strong evidence that justices rely more on the clerks in their first few years on the Court. If justices are adopting different approaches to opinion writing over their tenure, it does not manifest itself in the variability of their writing.

Our analysis also raises institutional questions about the Court, which has remained steadfastly lean over time. As in the 19th century, today's Court consists of only nine justices. Their primary support remains their law clerks; the only difference is that each justice has four rather than two justices. As the justices' caseload demands have steadily grown, it is not surprising that law clerks play an increasingly substantive role.

If our main concern is that justices delegate too much writing to their clerks, an obvious solution is to increase the number of justices. Court expansion would increase the number of justices to hear cases and write opinions. Article III of the U.S. Constitution is silent on the number of justices, so Court expansion would not require a constitutional amendment. This expansion, however, may be politically infeasible: intense confirmation hearings of any Court nominee appear to be the norm,[71] and Congress and the public vehemently opposed the last attempt to expand the Court.[72]

Another approach, recently proposed by scholars, is for the Court to hear cases in smaller panels (e.g., three justices) rather than *en banc*.[73] If we assume that the docket and the number of cases granted *certiorari* remained the same, each justice could presumably prepare fewer cases

---

[69] Mark Tushnet, Style and the Supreme Court's Educational Role in Government, 11 Constitutional Commentary 215, 222 (1994).

[70] *See* generally Garrow, *supra* note / /; and Stephen Calabresi and James Lindgren, *Term Limits for the Supreme Court: Life Tenure Reconsidered*, 29 Harv. J. Law & Pub. Pol. 770 (2006).

[71] *See* Elena Kagan, *Confirmation Messes Old and New*, 62 U. CHI. L. REV. 919, 921-30 (1995)

[72] *See* WILLIAM E. LEUCHTENBURG, THE SUPREME COURT REBORN: THE CONSTITUTIONAL REVOLUTION IN THE AGE OF ROOSEVELT 132-162 (1996)

[73] *See* Tracey E. George and Chris Guthrie, *Remaking the United States Supreme Court in the Courts' of Appeals Image*, 58 DUKE L. J. 1439 (2009); Tracey E. George and Chris Guthrie, "'the threes': Re-imagining Supreme Court Decisionmaking," 61 VAND. L. REV. 1825 (2008).

for oral argument and dedicate more time to writing opinions. The Court could make this change on its own, without any Congressional approval. Other countries have taken this exact approach, such as the Canadian Supreme Court.[74] It is unclear, however, whether the current Court would be willing to do so.

If, however, our concern is less about delegation, *per se*, and more about the implications of delegating so much to relatively inexperienced staff, then the Court could modify its clerk selection process. Rather than draw primarily from recent law graduates, the Court could hire more law clerks with greater legal experience. The intuition here is that if there are positive returns to experience in the study of law – implicit in our selection of justices – then it logically follows that these criteria are desirable attributes among law clerks. Some state supreme courts have incorporated this idea through the creation of permanent clerks.[75]

Conclusion

This article provides a statistical analysis of judicial opinions aimed at improving our understanding of judicial authorship. We show that it is possible to statistically evaluate the content of justices' opinions, and our results offer strong evidence that justices are increasingly relying on their clerks when writing opinions. Whether this trend is desirable is a separate, more involved discussion that goes beyond the scope of this article.

More broadly, this article provides what we hope is an important step toward bridging research between social science and traditional legal scholarship. The former focuses almost exclusively how judges and justices vote on individual cases and has relatively little to say substantively about legal doctrine, while the latter has much to say about legal doctrine but less systematically measuring how it develops or its subsequent effect. While this article examines judicial authorship, statistical analysis can allow us to examine important substantive issues, such as the evolution of legal doctrine across a federalist system.

[74] *See* Benjamin Alarie, Andrew Green, and Edward Iacobucci, *Is Bigger Always Better? On Optimal Panel Size, with Evidence from the Supreme Court of Canada,* working paper (2010) (available on file with the author).

[75] For example, the California Supreme Court and lower courts now have permanent law clerks. *See* Itir Yakar, "Unseen Staff Attorneys Anchor State's Top Court: Institution's System of Permanent Employees Means Workers Can Outlast the Justices," S.F. DAILY JOURNAL, 30 May 2006, 1.

Appendix

Table A1
Current and Departed Justices Pre-1941

Number	Justice	Appointing President	Year Joined Court	Year Left Court	Age at Retirement	Years on Court	Opinions	Average Word Length	V Score
42	Pierce Butler	Harding	1923	1939	73	16	321	1847	2.67
43	Louis Dembitz Brandeis	Wilson	1916	1939	83	23	446	1726	2.32
44	Benjamin Nathan Cardozo	Hoover	1932	1938	68	6	126	2699	2.30
45	George Sutherland	Harding	1922	1938	76	16	279	2271	2.34
46	Willis Van Devanter	Taft	1911	1937	78	26	344	2243	2.50
47	Oliver Wendell Holmes	Roosevelt, T.	1902	1932	91	30	861	1129	1.76
48	Edward Terry Sanford	Harding	1923	1930	65	7	128	1912	2.70
49	William Howard Taft	Harding	1921	1930	73	9	248	2759	2.82
50	Joseph McKenna	McKinley	1898	1925	82	27	620	2219	2.60
51	John Hessin Clarke	Wilson	1916	1922	65	6	121	1961	2.46
52	Mahlon Pitney	Taft	1912	1922	64	10	236	2899	2.90
53	William Rufus Day	Roosevelt, T.	1903	1922	73	19	407	2540	2.94
54	Edward Douglass White	Taft	1910	1921	76	11	469	4826	3.71
55	Joseph Rucker Lamar	Taft	1911	1916	59	5	107	1887	2.24
56	Charles Evans Hughes	Taft	1910	1916	54	6	390	2811	3.06
57	Horace Harmon Lurton	Taft	1910	1914	70	4	91	2970	2.86
58	John Marshall Harlan	Hayes	1877	1911	78	34	709	3377	3.44
59	William Henry Moody	Roosevelt, T.	1906	1910	57	4	62	2307	2.59
60	Edward Douglass White	Cleveland	1894	1910	65	16	628	3001	2.94
61	David Josiah Brewer	Harrison	1890	1910	73	20	507	2392	2.55
62	Melville Weston Fuller	Cleveland	1888	1910	77	22	714	2159	2.79
63	Rufus Wheeler Peckham	Cleveland	1896	1909	71	13	296	3199	3.33
64	Henry Billings Brown	Harrison	1891	1906	70	15	429	2884	3.30
65	George Shiras	Harrison	1892	1903	71	11	240	3066	3.15
66	Horace Gray	Arthur	1882	1902	74	20	429	2528	2.97
67	Stephen Johnson Field	Lincoln	1863	1897	81	34	521	2175	2.68
68	Howell Edmunds Jackson	Harrison	1893	1895	63	2	45	3592	3.58
69	Lucius Quintus C. Lamar	Cleveland	1888	1893	36	5	99	2915	3.31
70	Samuel Blatchford	Arthur	1882	1893	73	11	412	3343	4.73
71	Joseph P. Bradley	Grant	1870	1892	79	22	371	3072	3.22
72	Samuel Freeman Miller	Lincoln	1862	1890	74	28	581	2179	2.50
73	Stanley Matthews	Garfield	1881	1889	65	8	226	3262	3.31
74	Morrison Remick Waite	Grant	1874	1888	72	14	650	1399	2.36
75	William Burnham Woods	Hayes	1881	1887	63	6	159	2187	2.93
76	Ward Hunt	Grant	1873	1882	72	9	142	1643	2.67
77	Noah Haynes Swayne	Lincoln	1862	1881	77	19	315	1699	2.14
78	Nathan Clifford	Buchanan	1858	1881	77	23	370	3245	3.26
79	William Strong	Grant	1870	1880	72	10	231	2472	3.25
80	David Davis	Lincoln	1862	1877	62	15	182	1401	1.95

Table A1 (con't)
Current and Departed Justices Pre-1941

Number	Justice	Appointing President	Year Joined Court	Year Left Court	Age at Retirement	Years on Court	Opinions	Average Word Length	V Score
81	Salmon Portland Chase	Lincoln	1864	1873	65	9	114	1385	1.95
82	Samuel Nelson	Tyler	1845	1872	83	27	257	1784	2.29
83	Robert Cooper Grier	Polk	1846	1870	61	24	184	1642	2.20
84	James Moore Wayne	Jackson	1835	1867	90	32	119	4247	4.03
85	John Catron	Jackson	1837	1865	85	28	140	2078	2.23
86	Roger Brooke Taney	Jackson	1836	1864	81	28	237	2557	2.34
87	John Archibald Campbell	Pierce	1853	1861	58	8	88	1719	2.03
88	John McLean	Jackson	1830	1861	81	31	213	2395	2.61
89	Peter Vivian Daniel	Van Buren	1842	1860	68	18	80	3317	2.92
90	Benjamin Robbins Curtis	Fillmore	1851	1857	46	6	47	3273	2.71
91	John McKinley	Van Buren	1838	1852	68	14	11	3477	3.11
92	Levi Woodbury	Polk	1845	1851	57	6	41	3082	2.58
93	Joseph Story	Madison	1812	1845	77	33	228	2815	2.63
94	Henry Baldwin	Jackson	1830	1844	54	14	36	3801	2.74
95	Smith Thompson	Monroe	1823	1843	67	20	75	2947	2.70
96	Philip Pendleton Barbour	Jackson	1836	1841	55	5	17	4620	2.99
97	Gabriel Duvall	Madison	1811	1835	56	24	10	1174	2.16
98	John Marshall	Adams, John	1801	1835	80	34	385	2556	2.89
99	William Johnson	Jefferson	1804	1834	63	30	97	1884	2.00
100	Bushrod Washington	Adams, John	1799	1829	67	30	55	1981	2.27
101	Robert Trimble	Adams, J. Q.	1826	1828	43	2	14	2226	2.19
102	Thomas Todd	Jefferson	1807	1826	61	19	10	1090	1.96
103	Henry Brockholst Livingston	Jefferson	1807	1823	71	16	34	1548	1.79
104	Samuel Chase	Washington	1796	1811	70	15	114	1385	1.95
105	William Cushing	Washington	1790	1810	78	20	1	1561	
106	William Paterson	Washington	1793	1806	61	13	1	440	
107	Alfred Moore	Adams, John	1800	1804	49	4			
108	Oliver Ellsworth	Washington	1796	1800	55	4	2	1412	1.03
109	James Iredell	Washington	1790	1799	48	9			
110	James Wilson	Washington	1789	1798	56	9			
111	John Rutledge	Washington	1795	1795	56	0	1	727	
112	John Blair	Washington	1790	1795	63	5			
113	John Jay	Washington	1789	1795	50	6	1	988	
114	Thomas Johnson	Washington	1792	1793	61	1			
115	John Rutledge	Washington	1790	1791	52	1	65	4484	3.41

Note: in the early years on the Court (1790-1820), the identity of the author of an opinion was either the Chief Justice or not revealed. Accordingly, we could not produce a V score for many of these early justices.