STANFORD TECHNOLOGY LAW REVIEW
VOLUME 16, NUMBER 3 SPRING 2013

USING ALGORITHMIC ATTRIBUTION TECHNIQUES TO DETERMINE AUTHORSHIP IN UNSIGNED JUDICIAL OPINIONS

William Li,* Pablo Azar,* David Larochelle,* Phil Hill,* James Cox,* Robert C. Berwick,* & Andrew W. Lo* †

CITE AS: 16 STAN. TECH. L. REV. 503 (2013)
http://stlr.stanford.edu/pdf/algorithmicattribution.pdf

ABSTRACT

This Article proposes a novel and provocative analysis of judicial opinions that are published without indicating individual authorship. Our approach provides an unbiased, quantitative, and computer scientific answer to a problem that has long plagued legal commentators.

---

* William Li is a PhD student in the Computer Science and Artificial Intelligence Laboratory (CSAIL) and a 2012 graduate of the Technology and Policy Program at the Massachusetts Institute of Technology (MIT).
* Pablo Azar is a PhD student in the Computer Science and Artificial Intelligence Laboratory (CSAIL) at the Massachusetts Institute of Technology (MIT).
* David Larochelle is an engineer at the Berkman Center for Internet & Society at Harvard University.
* Phil Hill is a Fellow at the Berkman Center for Internet & Society at Harvard University and a 2013 J.D. Candidate at Harvard Law School.
* James Cox was an associate with Jenner & Block LLP during drafting of this Article, and currently serves as an attorney for the United States government.
* Robert C. Berwick is Professor of Computational Linguistics and Computer Science and Engineering in the Departments of Electrical Engineering and Computer Science and Brain and Cognitive Sciences, MIT.
* Andrew W. Lo is the Charles E. and Susan T. Harris Professor at the MIT Sloan School of Management, Principal Investigator in the Computer Science and Artificial Intelligence Laboratory (CSAIL) at the Massachusetts Institute of Technology (MIT), and a joint faculty in the MIT Electrical Engineering and Computer Science Department.
† We thank John Cox at MIT, Andy Sellars and Ryan Budish at the Berkman Center, and Philip C. Berwick at the Washington University in St. Louis Law School for their invaluable feedback, and Jayna Cummings for editorial assistance.

United States courts publish a shocking number of judicial opinions without divulging the author. Per curiam opinions, as traditionally and popularly conceived, are a means of quickly deciding uncontroversial cases in which all judges or justices are in agreement. Today, however, unattributed per curiam opinions often dispose of highly controversial issues, frequently over significant disagreement within the court. Obscuring authorship removes the sense of accountability for each decision's outcome and the reasoning that led to it. Anonymity also makes it more difficult for scholars, historians, practitioners, political commentators, and—in the thirty-nine states with elected judges and justices—the electorate, to glean valuable information about legal decision-makers and the way they make their decisions. The value of determining authorship for unsigned opinions has long been recognized but, until now, the methods of doing so have been cumbersome, imprecise, and altogether unsatisfactory.

Our work uses natural language processing to predict authorship of judicial opinions that are unsigned or whose attribution is disputed. Using a dataset of Supreme Court opinions with known authorship, we identify key words and phrases that can, to a high degree of accuracy, predict authorship. Thus, our method makes accessible an important class of cases heretofore inaccessible. For illustrative purposes, we explain our process as applied to the Obamacare *decision, in which the authorship of a joint dissent was subject to significant popular speculation. We conclude with a chart predicting the author of every unsigned per curiam opinion during the Roberts Court.*

INTRODUCTION

U.S. courts publish a shocking number of opinions without divulging the author. Unsigned per curiam opinions, as traditionally and popularly conceived, are a means of quickly deciding uncontroversial cases in which all judges or justices are in agreement. Today, however, unsigned per curiam opinions often dispose of highly controversial issues, frequently over significant disagreement within a court. Obscuring authorship removes the sense of accountability for each decision's outcome and the reasoning that led to it. Anonymity also makes it more difficult for scholars, historians, practitioners, political commentators, and—where applicable—the electorate, to glean valuable information about legal decision-makers and the way they make their decisions. The value of determining authorship for unsigned opinions has long been recognized but, until now, the methods of doing so have been cumbersome, imprecise, and altogether unsatisfactory. Currently, to obtain information on how decisions were made and authored, the public relies on anecdotal evidence from clerks, legal observers, and occasional comments by judges and justices themselves.

Given the importance of unsigned opinions and the large corpus of signed judicial writings, we demonstrate that novel computational tools can add quantitative, non-partisan insight into judicial opinion authorship. Our work uses statistical data mining and machine learning algorithms to predict authorship of judicial opinions that are unsigned or whose attribution is disputed. Using a dataset of Supreme Court opinions with known authorship, we identify key words and phrases that can, to a high degree of accuracy, predict authorship using only the text from a judicial opinion (with obvious identifying markers removed). After training "writing style models" for the different justices under consideration, we can predict the author of an unsigned opinion by analyzing only that unsigned opinion's text. Our method provides insight into which authors were most influential in writing the published opinion, thereby giving interested parties access to an important class of cases heretofore inaccessible.

Part I summarizes the historical context of unsigned per curiam opinions, criticisms of the practice, and compares our attribution solution to other approaches. To illustrate our method, Parts II–IV describe the process of determining authorship in a recent, high profile Supreme Court case in which the author of the dissenting opinion was subject to much popular speculation. Part II describes this illustrative test case. Part III describes the experimental setup and Part IV explains the results. Part V provides the results of applying our process to every unsigned per curiam opinion of the Roberts Court.

I. UNSIGNED OPINIONS

Over the last 150 years, there has been an astonishing number of court decisions issued without attribution. Unsigned opinions have been the subject of great controversy since their inception and present many problems today.

Although unsigned opinions appear in all appellate courts—federal or state—the example set by the Supreme Court is particularly illustrative. Part I.A. provides historical context for unsigned per curiam opinions. Part I.B. summarizes some of the problems that have been associated with judicial anonymity. Part I.C. evaluates some of the current methods used to determine authorship. Part I.D. briefly describes how our method provides a better means of determining authorship.

A. *Historical Context of Unsigned Opinions*

The Supreme Court's attribution practices have a long and colorful history.[1] The Court has delivered opinions in one of four ways.[2] First, in the early days, the Court would issue decisions "seriatim," whereby the Justices wrote separate opinions that were published in order of seniority, and were sometimes followed by a summary order "By the Court" with the overall disposition.[3] Second, for uncontroversial and unanimous decisions, the Court would deliver an opinion under the heading "By the Court."[4] Third, the Court would deliver a single opinion under the name of the Chief Justice, while indicating that he was speaking "for the Court."[5] Finally, the Court would issue a majority opinion with justices writing separately as they desired.[6]

The fourth attribution option—a majority opinion accompanied by separate dissents and concurrences—has become the familiar means of delivering opinions.[7] However, during the Marshall era, the Chief Justice chose the third option in an effort to enhance the Court's image of solidarity and authority.[8] Even when the justices disagreed, Marshall insisted that the Court issue only one opinion in his name, even when he disagreed in the judgment.[9] This practice gradually gave way and by 1832, all members of the Court had written separately at least once. By the time Marshall died in 1835, the practice of writing separate opinions was solidified.[10]

Throughout these shifts in the Court's attributional philosophy, the "per curiam" decision—in which an opinion states the ostensible opinion of "the Court" rather than any particular justice(s)—has remained a viable option.

---

1. For a more robust history, see generally John P. Kelsh, *The Opinion Practices of the United States Supreme Court 1790–1945*, 77 WASH. U. L.Q. 137 (1999); James Markham, Note, *Against Individually Signed Judicial Opinions*, 56 DUKE L.J. 923, 928 (2006).
2. Markham, *supra* note 1, at 928.
3. *Id.*
4. *Id.*
5. *Id.*
6. *Id.*
7. *Id.* at 929.
8. *Id.*
9. *Id.*
10. *Id.*

However, such decisions have not been confined to uncontroversial or unanimous topics as they were in the early days of the Court. The classic occasion for a per curiam decision is when the law is so clear that the justices are unanimous and the issue does not merit the time necessary to craft a detailed opinion.[11]

However, with great frequency today's per curiam opinions are neither unanimous nor uncontroversial. A 1992 study found that only 44% of per curiam opinions were unanimous.[12] For the other 56% of per curiam opinions, there are two ineluctable conclusions: (1) despite the label, an ostensibly "per curiam" opinion *cannot* speak "for the Court" as a whole, and (2) there is at least some controversy to the disposition. Some of the most important and controversial cases in our nation's history came through badly divided per curiam opinions.[13] Such decisions include invalidating the death penalty in *Furman v. Georgia* (five concurrences, four dissents),[14] dealing with campaign finance reform in *Buckley v. Valeo* (involving a 137-page per curiam with five opinions concurring in part and dissenting in part),[15] resolving the Pentagon Papers case (six concurrences, three dissents),[16] and ending the presidential election of 2000 in *Bush v. Gore* (one concurrence, four dissents),[17] to name a few.

With these qualitative concerns in mind, the number of per curiam opinions is alarming. The Warren Court used per curiam opinions 28.7% of the time, the Berger Court 17.7%, the Rehnquist Court 10.3%, and the Roberts Court 13.3%.[18] In 2011, the federal courts of appeal issued per curiam opinions 7.6% of the time, with significant variation across circuits.[19] Whereas the D.C. Circuit relied on per curiam opinions only 0.3% of the time, the Fifth Circuit used them 15.9% of the time.[20] The problem is direr in some state courts, where per curiam opinions constitute *more than half* of an elected court's decisions.[21]

---

11. *See, e.g.*, *id.* at 934; Ira P. Robbins, *Hiding Behind the Cloak of Invisibility: The Supreme Court and Per Curiam Opinions*, 86 TUL. L. REV. 1197, 1200-02 (2012); Laura Krugman Ray, *The Road to* Bush v. Gore*: The History of The Supreme Court's Use of the Per Curiam Opinion*, 79 NEB. L. REV. 517, 521-24 (2000); Stephen L. Wasby et al., *The Per Curiam Opinion: Its Nature and Functions*, 76 JUDICATURE 29, 30 (1992).

12. Wasby et al., *supra* note 11, at 35.

13. *See* Michael C. Gizzi & Stephen L. Wasby, *Per Curiams Revisited: Assessing the Unsigned Opinion*, 96 JUDICATURE 110, 113 (2012).

14. 408 U.S. 238 (1972).

15. 424 U.S. 1 (1976).

16. 403 U.S. 713 (1971).

17. 531 U.S. 98 (2000).

18. Gizzi & Wasby, *supra* note 13, at 111.

19. *Id.* at 114.

20. *Id.* at 115.

21. *In the Shadows: A Look into the Texas Supreme Court's Overuse of Anonymous Opinions*, at 1, TEXAS WATCH (May 2008), *available at* http://www.texaswatch.org/wordpress/wp-content/uploads/2009/12/PerCuriamReportFinal.pdf.

B. *Problems with Unsigned Opinions*

The previous section highlighted the quantity and quality of cases using unsigned opinions, but what is the harm? The most compelling complaints focus on the theme of accountability: poor quality of opinions, evasion of difficult issues, lack of transparency to the public, and the like.[22]

Critics voicing the accountability concern are numerous and frequently high profile. In response to Chief Justice Marshall's edict that the Supreme Court issue a single opinion in his name, Thomas Jefferson wrote that "secret, unanimous opinions" written on behalf of the Court would undermine judicial accountability.[23] When "nobody knows what opinion any individual member gave in any case, nor even that he who delivers the opinion concurred in it himself, [a justice's reputation] is shielded completely."[24] Jefferson disapproved of opinions reached by justices "huddled up in a conclave, perhaps by a majority of one, delivered as if unanimous, and with the silent acquiescence of lazy or timid associates, by a crafty chief judge, who sophisticates the law to his own mind, by the turn of his own reasoning."[25] Per curiam opinions, according to Jefferson, are "certainly convenient for the lazy, the modest, and the incompetent."[26]

Jefferson's disapproval and call for accountability has echoed since. President Madison called for a return to seriatim opinions "so that Republican judges could record their position on the issues."[27] When she was a circuit judge, Justice Ginsburg wrote, "Public accountability through the disclosure of votes and opinion authors puts the judge's conscience on the line."[28] She further noted, "Judges generally do not labor over unpublished judgments and memoranda, or even per curiam opinions, with the same intensity they devote to signed opinions."[29] Approvingly quoting another commentator, Justice Ginsburg wrote, "[W]hen anonymity of pronouncement is combined with security in office, it is all too easy, for the politically insulated officials to lapse into arrogant ipse dixits."[30] Judge Richard Posner agreed, asserting that signed opinions elicit the greatest effort from judges and make "the threat of searing

22. *See generally* Robbins, *supra* note 11. Other critiques include stunting the development of the law by reducing the ability to analyze a judge or justice's jurisprudence and put to use any lessons derived from such analysis. *See id.* 1224-41.

23. Markham, *supra* note 2, at 930 (quoting Letter from President Thomas Jefferson to Justice William Johnson (Oct. 27, 1820)).

24. *Id.*

25. *Id.* (quoting Letter from Thomas Jefferson to Thomas Ritchie (Dec. 25, 1820)).

26. *See* Kelsh, *supra* note 1, at 145-46 (citing Letter from Thomas Jefferson to William Johnson (Oct. 27, 1822)).

27. *Id.*

28. Ruth Bader Ginsburg, *Remarks on Writing Separately*, 65 WASH. L. REV. 133, 140 (1990).

29. *Id.* at 139.

30. *Id.*

professional criticism an effective check on irresponsible judicial actions."[31] Discussing the Court's decision in *Bush v. Gore*, one commentator noted that per curiams are convenient tools in controversial cases because "[w]ith no Justice signing the opinion, there [is] no individual to be blamed for evading the tough questions."[32]

These accountability criticisms have even more force when the judges and justices are elected officials serving terms of office rather than appointed judges and justices serving for life or good behavior. Thirty-nine states (78%) require judges to run for election or win periodic retention votes.[33] The electorates in these states need recorded votes and opinions to evaluate their respective judges and justices, but unsigned opinions reduce access to this vital information.

For example, Texas is one state in which judges are elected by and accountable to voters. During the 2006–2007 term, an astounding 57% of the opinions issued by the Supreme Court of Texas were unsigned per curiams.[34] Over a ten-year period, per curiam opinions constituted 40% of the opinions issued by the Supreme Court of Texas.[35] One commentator puts the problem nicely:

> When a judge signs his name to an opinion he has written, he accepts responsibility for the decision and the logic used in reaching it. Whether the opinion is a stellar example of judicial wisdom or a blatant abuse of judicial authority, the author is accountable because his identity is known. Any judge who disagrees with an authored opinion must write or join a dissent, and thus that judge's position is known as well, and he is equally accountable.
>
> When a court releases a per curiam opinion, however, no judge accepts responsibility for the opinion, and no judge can be held accountable for it. The public does not know if all judges agreed with the holding. Judges who can hide behind this anonymity may not have an incentive to reach the legally correct conclusion or to justify the conclusion they do reach.[36]

This same commentator goes on to list controversial per curiam opinions in the Texas Supreme Court and analyzes campaign contributions from parties who appeared before the court in such cases.[37] This example highlights one of the most extreme consequences of judicial unaccountability.

C. *Solving Attributional Questions the Old-Fashioned Way*

Scholars and historians have long been skeptical of unsigned opinions and have sounded numerous calls for research identifying the authors of unsigned

---

31. RICHARD A. POSNER, THE FEDERAL COURTS: CHALLENGE AND REFORM 349 (1999).
32. *See* Ray, *supra* note 11, at 521-22.
33. *See* Robbins, *supra* note 11, at 1221.
34. *See In the Shadows*, *supra* note 20.
35. *Id.*
36. *Id.*
37. *See generally id.*

opinions. Until now, the methodologies recommended and employed have been decidedly "old school."

A 2012 article charting the use of per curiam decisions of the Supreme Court suggested some methods for determining authorship.[38] First, the article recommends narrowing the possibilities by ruling out authors of separately signed opinions.[39] However, this method will frequently yield incomplete results. At best, it narrows the possibilities to the handful of justices who did not write a separate opinion. However, this approach does not rule out the possibility that one of the justices authored *both* a signed opinion and the unsigned per curiam. Moreover, this step is useless when there is no separately written opinion.

The article also recommends culling the files of retired Justices like Blackmun and White, which are available in the Manuscript Division of the Library of Congress.[40] Apart from practical access difficulties, the files are incomplete records of communications between the justices and other confidants. Some useful narratives may be pieced together with effort, but these records are still likely to prove incomplete and unsatisfactory. Furthermore, this information only becomes available after a justice retires, which will significantly delay authorship investigations in the vast majority of cases. Even after those files become available, there is no guarantee that the information found therein will be of any use.

Investigations into unsigned opinions from federal appellate and state courts will experience additional problems.[41] Unlike the Supreme Court, these courts are less likely to maintain robust records of any behind-the-scenes happenings. Moreover, such opinions have less practical significance relative to U.S. Supreme Court opinions. In turn, this consideration may imply that historians and other parties have less motivation to investigate how these courts arrived at their decisions and to publish such findings for the benefit of further research.

D. *Solving Attributional Questions Algorithmically*

Our approach, involving algorithmic natural language analysis, presents several advantages over the aforementioned approaches to determining authorship. First, access is practically a non-issue. Using only public domain opinions of known authorship, we can create a dataset from which we can analyze the natural language of any given opinion. A sufficient quantity of opinions for the dataset is often freely available through resources like the Cornell Legal Information Institute (LII).[42]

---

38. *See* Gizzi & Wasby, *supra* note 13, at 116.
39. *Id.*
40. *Id.* at 116-17.
41. *Id.* at 118.
42. *Supreme Court Collection*, LEGAL INFORMATION INSTITUTE, http://www.law.

We need not access the Library of Congress or put together an incomplete puzzle from the files of retired justices. For our system, we need only an unsigned opinion and a sufficiently large bank of signed opinions from the potential authors. Part V lays out the results from applying our algorithm to every unsigned opinion of the Roberts Court. But for illustrative purposes, the next few Parts describe our approach when used on a high-profile test case with an opinion whose authorship was hotly contested.

II. TEST CASE: *OBAMACARE*

In June 2012, the Supreme Court issued its ruling in *National Federation of Independent Business v. Sebelius*,[43] which largely upheld the 2010 Patient Protection and Affordable Care Act (PPACA). This highly controversial decision contained what was originally an unsigned dissenting opinion, the authorship of which was a popularly debated issue.

The *Sebelius* decision was surprising because Chief Justice John Roberts gave the deciding vote, siding with the more liberal justices.[44] The Chief Justice rejected the government's argument that Congress was authorized to enact PPACA's individual insurance coverage mandate under the Commerce Clause, but accepted the government's alternative position that the mandate was authorized by Congress's power to enact taxes.[45] Together with the liberal justices, who would have accepted both government arguments, the Chief Justice provided the necessary fifth vote to uphold the law.[46]

Many experts had predicted that (1) the Court would overturn PPACA[47] and (2) the pivotal vote would come from Justice Anthony Kennedy.[48] After the decision, there was speculation that the Chief Justice had switched sides between the time that the case was heard and the time the decision was announced.[49] There was further speculation that the formerly unsigned dissent

---

cornell.edu/supct/ (last visited Feb. 11, 2013).

43. 132 S. Ct. 2566 (2011).

44. *See, e.g.*, John T. Bennett, *Law of the Land: Supreme Court Upholds* 'Obamacare,' U.S. NEWS (June 28, 2012), *available at* http://www.usnews.com/news/articles/2012/06/28/law-of-the-land-supreme-court-upholds-obamacare.

45. 132 S. Ct. at 2587.

46. *See* Amy Davidson, *Roberts the Swing Vote: Court Upholds Most of Health Care*, The NEW YORKER (Jun. 28, 2012), *available at* http://www.newyorker.com/online/blogs/closeread/2012/06/roberts-the-swing-vote-court-upholds-most-of-health-care.html; Adam Winkler, *The Roberts Court is born*, SCOTUSblog (Jun. 28, 2012, 12:01 PM), http://www.scotusblog.com/2012/06/the-roberts-court-is-born/.

47. *See, e.g.*, Peter Ferrara, *Why the Supreme Court Will Strike Down All of* Obamacare, FORBES (Apr. 5, 2012), *available at* http://www.forbes.com/sites/peterferrara/2012/04/05/why-the-supreme-court-will-strike-down-all-of-obamacare/.

48. *See, e.g.*, Peter J. Boyer, *Reading Justice Anthony Kennedy's Leanings on* Obamacare, THE DAILY BEAST (Apr. 2, 2012), http://www.thedailybeast.com/newsweek/2012/04/01/reading-justice-anthony-kennedy-s-leanings-on-obamacare.html.

49. *See, e.g.*, Sabrina Siddiqui, *John Roberts' Switch on* Obamacare *Sparks Fascination with Supreme Court, Possible Leaks*, HUFFINGTON POST (July 3, 2012),

(later attributed to Justices Kennedy, Scalia, Thomas, and Alito)[50] had originally been a majority opinion authored by Chief Justice Roberts.[51]

The following pieces of evidence have been offered to support this hypothesis. First, the opening section of the joint dissent authored by Kennedy et al. (the "joint dissent") never mentions the Court's majority opinion to uphold the PPACA, written by Chief Justice Roberts.[52] Typically, dissenting and concurring opinions will highlight in the first few paragraphs the reason for authoring a separate opinion, as indeed Justice Ginsburg's opinion[53] and Justice Thomas's opinion[54] do. Instead, the joint dissent only contains arguments against points made by the government attorneys defending PPACA and Justice Ginsburg's opinion. In this respect, the joint dissent reads more like a majority opinion (with a corresponding dissent by Justice Ginsburg), rather than a dissent arguing against Chief Justice Robert's opinion.

Second, whereas a typical majority opinion will refer to the decision-maker as the Court and describe the majority justices using the collective pronoun "we," indicating solidarity, dissents and concurrences typically refer to themselves individually using less inclusive pronouns like "I." True to form, Chief Justice Roberts' majority opinion follows this pattern,[55] as does Justice Ginsburg's opinion (which is joined by Justice Sotomayor),[56] and Justice Thomas's opinion.[57] The joint dissent, however, does not follow the pattern.[58]

Third, although there are two dissenting opinions in this case—the joint dissent and another authored by Justice Thomas alone—the joint dissent refers to Justice Ginsburg's opinion concurring in part and dissenting in part in the following way: "*The* dissent claims that we 'fai[l] to explain why the individual mandate threatens our constitutional order.' *Ante*, at 2627. But we have done so."[59] It is peculiar that this joint dissent does not acknowledge itself as one of

---

http://www.huffingtonpost.com/2012/07/02/justice-roberts-obamacare-supreme-court-leaks_n_1644864.html; Paul Campos, *Did John Roberts Switch His Vote?*, SALON.COM (June 28, 2012), http://www.salon.com/2012/06/28/did_john_roberts_switch_his_vote/.

50. 132 S. Ct. 2642.

51. *See, e.g.*, Avik Roy, *The Inside Story on How Roberts Changed His Supreme Court Vote on* Obamacare, FORBES (July 1, 2012), *available at* http://www.forbes.com/sites/aroy/2012/07/01/the-supreme-courts-john-roberts-changed-his-obamacare-vote-in-may/.

52. *See* 132 S. Ct. 2642-44.

53. *Id.* at 2602 (Ginsburg, J., concurring in part and dissenting in part).

54. *Id.* at 2677 (Thomas, J., dissenting).

55. *E.g.*, *id.* at 2576 ("Today *we* resolve . . . . *We* do not consider . . . . *We* ask only . . . ." (emphasis added)).

56. *Id.* at 2609 (Ginsburg, J., concurring in part and dissenting in part) ("*I* agree with . . . *I* therefore join . . . [H]owever, *I* would hold . . . . (emphasis added)). Note also that Justice Ginsburg uses first person singular pronouns despite the fact that Justices Sotomayor, Breyer, and Kagan joined in all or part of the Justice Ginsburg's opinion. *Id.*

57. *Id.* at 2677 (Thomas, J., dissenting) ("*I* dissent . . . but *I* write separately to . . . . *I* adhere to *my* view" (emphasis added)).

58. *Id.* at 2432 (Scalia, Kennedy, Thomas, Alito, Js., dissenting) (emphasis added) ("*We* conclude . . . .").

59. *Id.* at 2659 (Scalia, Kennedy, Thomas, Alito, Js., dissenting) (emphasis added).

the two opinions dissenting in full. It is even more peculiar that the joint dissent calls Justice Ginsburg's opinion "*the* dissent" when her opinion is, in fact, only partially dissenting. As a practical matter, it would appear that either the joint dissent or Justice Thomas' dissent is more deserving of being dubbed "*the* dissent." This sentence would make more sense as a majority opinion critiquing a sole dissenting opinion (on the assumption that, in this counterfactual scenario, Justice Thomas would have joined the counterfactual majority or at least changed his dissent to a concurrence).

Fourth, Justice Ginsburg provided the following criticism of Chief Justice Roberts's reasoning in the majority opinion:

> In failing to explain why the individual mandate threatens our constitutional order, THE CHIEF JUSTICE disserves future courts. How is a judge to decide, when ruling on the constitutionality of a federal statute, whether Congress employed an independent power, *ante*, at 2591, or merely a derivative one, *ante*, at 2592. Whether the power used is substantive, *ante*, at 2592, or just incidental, *ante*, at 2592? The instruction THE CHIEF JUSTICE, in effect, provides lower courts: You will know it when you see it.[60]

There is a direct response to this argument, but it appears in the joint dissent, not Chief Justice Roberts' majority opinion.[61]

An alternate hypothesis is that the joint dissent was actually written mostly by Justices Kennedy and Scalia, as argued by a detailed news article with sources allegedly close to the Supreme Court.[62] This article also gives an explanation as to why the joint dissent does not engage Justice Roberts' majority opinion:

> The majority decisions were due on June 1, and the dissenters set about writing a response, due on June 15. The sources say they divided up parts of the opinion, with Kennedy and Scalia doing the bulk of the writing. The two sources say suggestions that parts of the dissent were originally Roberts' actual majority decision for the court are inaccurate, and that the dissent was a true joint effort.
>
> The fact that the joint dissent doesn't mention Roberts' majority was not a sign of sloppiness, the sources said, but instead was a signal the conservatives no longer wished to engage in debate with him.[63]

A further interview with Justice Ginsburg suggests that she wrote her own dissent early on, believing that the Chief Justice would strike down the individual mandate:

> Ginsburg quickly began drafting the dissenting statement on that issue, portions of which she read from the bench on the day the ruling was announced. "I had a draft of the dissent before the chief circulated his opinion because I knew it would be impossible to do" as the term went into the final

60. *Id.* at 2627-28 (Ginsburg, J., concurring in part and dissenting in part).

61. *Id.* at 2649 (Scalia, Kennedy, Thomas, Alito, Js., dissenting).

62. Jan Crawford, *Roberts Switched Views to Uphold Health Care Law*, CBS NEWS, (July 1, 2012), *available at* http://www.cbsnews.com/8301-3460_162-57464549/roberts-switched-views-to-uphold-health-care-law/.

63. *Id.*

month of June and several cases culminated.[64]

Ultimately, the evidence is mixed as to whether both the majority opinion and the joint dissent were authored by the Chief Justice. Applying authorship attribution techniques, we aspire to determine quantitatively which of these hypotheses is more plausible. Our model assigns the highest probability to Justices Scalia and Kennedy, not Chief Justice Roberts, as the author of the dissenting opinion, which supports the "Crawford" theory of authorship.[65]

III. EXPERIMENTAL SETUP

A. *Experimental Questions*

To solve this attribution question, we focused on whether features of each justice's writing styles could be used to predict which justice authored which opinion. The specific questions that we sought to answer through our experiments were the following:

1. Can statistical authorship attribution methods accurately predict which of the Supreme Court justices authored a given opinion?
2. What are the words and stylistic features that most distinguish different Supreme Court justices, and what do they reveal about the writing styles of different justices?
3. Which author(s) does the model predict for the majority and dissenting opinions written in *Sebelius*?

Our work is part of the growing literature on applying algorithmic natural language processing tools to legal opinions. Recent work has explored the evolution of language in Supreme Court texts over time,[66] the conversational dynamics of Supreme Court oral arguments,[67] and the role of law clerks in the opinion-writing process.[68] Our work applies an analogous quantitative approach to investigate the authorship of unsigned opinions and opinions of controversial attribution, leveraging advances in computational power and machine learning algorithms to infer authorship with high accuracy.

---

64. Joan Biskupic, *Exclusive: Justice Ginsburg Shrugs Off Rib Injury*, REUTERS (Aug. 8, 2012), *available at* http://www.reuters.com/article/2012/08/09/ us-usa-court-ginsburg-idUSBRE87801920120809.

65. *See* Crawford, *supra* note 62.

66. David Katz et al., *Legal n-grams? A Simple Approach to Track the Evolution of Legal Language*, in PROCEEDINGS OF THE 24TH INTERNATIONAL CONFERENCE ON LEGAL KNOWLEDGE AND INFORMATION (2011).

67. Timothy Hawes et al., *Elements of a Computational Model for Multi-Party Discourse: The Turn-Taking Behavior of Supreme Court Justices*, 60(8) J. AM. SOC'Y INFO. SCIENCE & TECH. 1607 (2009).

68. Jeffrey S. Rosenthal & Albert H. Yoon, *Detecting Multiple Authorship of United States Supreme Court Legal Decisions Using Function Words*, 5(1) ANNALS OF APPLIED STATISTICS 283 (2011).

B. *Data Preparation*

We obtained texts of Supreme Court opinions from the Cornell Legal Information Institute (LII).[69] From this source, we downloaded all Supreme Court decisions written by the nine justices currently sitting on the Court who have served during the tenure of Chief Justice John Roberts, i.e., from 2005 to 2011. For each case, we extracted the majority and dissenting opinions if they existed, keeping track of their respective authors. We masked the surnames of the justices themselves and years, to avoid simply using name or year information to identify the author. Concurrences to either the majority or dissenting opinion were not included in our corpus of possible cases, as our focus was on predicting authorship of majority and dissenting opinions.

Using these criteria, our dataset consists of 568 opinions. In addition to these 568 opinions, using the same protocol, we obtained the majority (signed by Chief Justice Roberts) and dissenting (signed by Justices Scalia, Kennedy, Thomas, and Alito) opinions of the *Sebelius* decision and 65 per curiam decisions by the Roberts Court (until November 2012).

C. *Machine Learning System Overview*

Our machine learning approach follows the paradigm of "supervised learning": our algorithms identify characteristics (called "features") of each justice's writing style from opinions known to be authored by him or her. These characteristics are encoded in a statistical prediction model that describes the writing styles of the justices under consideration. Given a new opinion with an unknown author, the model predicts which justice wrote the opinion. It is worth noting that "supervised learning-based authorship attribution" has been applied to a wide range of literary, historical, and contemporary domains,[70] including studies on the Federalist Papers,[71] Shakespeare's plays,[72] and more recently, on large numbers of authors in online blogs or forums.[73]

Our system builds upon some early work in judicial authorship, but with a greater focus on predicting who authored a particular opinion. In the

---

69. *See supra* note 42.

70. For a detailed review of classification techniques, useful features, and application areas see Moshe Koppel et al., *Computational Methods in Authorship Attribution*, 60(1) J. AM. SOC'Y INFO. SCIENCE & TECH. 9 (2009), *available at* http://dx.doi.org/10.1002/asi.v60:1; Efstathios Stamatatos, *A Survey of Modern Authorship Attribution Methods.,* 60(3) J. AM. SOC'Y INFO. SCIENCE & TECH. 538 (2009), *available at* http://dx.doi.org/10.1002/asi.21001.

71. *See, e.g.*, Frederick Mosteller & David Wallace, INFERENCE AND DISPUTED AUTHORSHIP: THE FEDERALIST (1964).

72. *See, e.g.*, Thomas V.N. Merriam & Robert A.J. Matthews, *Neural Computation in Stylometry II: An Application to the Works of Shakespeare and Marlowe*, 9(1) LITERARY & LINGUISTIC COMPUTING 1 (1994).

73. *See, e.g.*, Moshe Koppel et al., *Authorship Attribution with Thousands of Candidate Authors*, *in* PROCEEDINGS OF THE 29TH ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL 659 (2006).

aforementioned studies on the role of clerks in opinion-writing, researchers found that they could differentiate between pairs of Supreme Court justices using just 63 function words.[74] Our work leverages a much larger number of words and word phrases (about 100 times as many) culled from the opinions themselves, a process that is feasible and inexpensive on today's computers. In doing so, we are able to handle the problem of accurately predicting which of the nine justices wrote an unsigned or controversial opinion.

Within this framework, our corpus of opinions is divided into three datasets:

1. "Training set": Of the 568 signed opinions, we take 451 of them (80%) to "train" the authorship attribution system. The machine learning algorithms, described further below, are given both the text and the author of each of these opinions and learn the parameters of this system from this training data.
2. "Validation set": The remaining 117 (20%) signed opinions are deliberately excluded from the training process, and the authorship attribution system is used to "predict" the authors of these cases. The trained system is given only the text of these opinions and asked to provide an authorship prediction. Given that the author is known in these cases, the prediction has little scholarly value in itself; however, the performance of the system on these 117 cases, which can be measured by comparing the prediction to the actual author, provides some indication of the system's predictive capabilities.
3. "Test set": The opinions of the *Sebelius* decision and the 65 per curiam opinions of the Roberts Court form the final test set. Similar to the validation set, these cases are excluded from the training process. The results of this analysis are shown in Parts IV and V.

For our Supreme Court authorship attribution task, there are nine justices and thus the model must accurately choose among nine choices for each opinion. Along with high classification accuracy, we identified the following desiderata (in consultation with a practicing attorney familiar with Supreme Court cases and customs) for our classification scheme: (1) the features should be intuitive and easy to understand; (2) the prediction should have a confidence score for the correctness of the predicted justice; (3) the prediction should produce a meaningful probability distribution over the nine justices; and (4) the predicted author should be the justice with the highest probability.

D. *Design of Authorship Attribution System*

Building a statistical authorship attribution model requires three main design decisions: (1) how to represent each judicial opinion, (2) which statistical machine learning model to use, and (3) how to select features. This

74. *See* Rosenthal & Yoon, *supra* note 68.

section describes how we made these decisions, guided by established practices in constructing authorship attribution systems, quantitative experiments (detailed in the next section) to validate our choices, and the specific questions we sought to answer with this model.

1. *Document Representation*

To build the statistical authorship attribution model, the opinions must be characterized in terms of numerical features. Human experts might examine each justice's vocabulary richness, grammatical patterns, opinion length, or other writing style characteristics to try to distinguish between them. In our method, we use straightforward features that serve as a proxy for these intuitions: the presence of one-, two-, and three-word sequences (known in natural language processing as unigrams, bigrams, and trigrams, respectively; "*n*-grams" is the term for any sequence of *n* words) in a document. Table 1 illustrates how a single sentence from the majority opinion of the *Sebelius* decision can be described by these word sequences. Applied over an entire written opinion, such features encode information about vocabulary, syntax (such as the use of "however" in the middle of sentence), and subject matter. We do not eliminate capitalization or punctuation marks, which may also be indicative of writing style; for example, "Namely" is a different feature than "namely.". In addition, we did not discard punctuation immediately following words because these characteristics might differentiate the justices' writing styles. These *n*-gram features have been effective in a wide range of authorship attribution efforts.[75]

TABLE 1: EXAMPLE OF SENTENCE DECOMPOSED INTO UNIGRAMS, BIGRAMS, AND TRIGRAMS

Full sentence	It does not, however, control whether an exaction is within Congress's power to tax.
Unigrams	"It"; "does"; "not,"; "however,"; "control"; "whether"; "an"; "exaction"; "is"; "within"; "Congress's"; "power"; "to"; "tax."
Bigrams	"It does"; "does not,"; "not, however,"; "however, control"; "control whether"; "whether an"; "an exaction"; "exaction is"; "is within"; "within Congress's"; "Congress's power"; "power to"; "to tax."
Trigrams	"It does not"; "does not, however"; "not, however, control"; "however, control whether"; "control whether an"; "whether an exaction"; "an exaction is"; "exaction is within"; "is within Congress's"; "within Congress's power"; "Congress's power to"; "power to tax."

Experimental Evaluation: We evaluated the authorship attribution model using four different sets of features: (1) unigrams only; (2) bigrams only; (3)

75. *See* Koppel et al., *supra* note 70.

trigrams only; and (4) the combination of unigrams, bigrams, and trigrams. A comparison of model performance across these sets of features allows us to measure the incremental value of longer sequences of words. The results are presented in Table 3 in the next Part.

2. *Model Selection*

Given our goal of producing a meaningful probability distribution of authorship, we trained a maximum entropy (MaxEnt) statistical model—which enjoys widespread use in text classification tasks—for our authorship attribution system.[76] Specifically, we designed our model to compute the following probability:

$$P(y_i|x) = \frac{\exp\left(\theta \cdot \phi(x, y_i)\right)}{\sum_{k=1}^{9} \exp\left(\theta \cdot \phi(x, y_k)\right)}$$

where:

x : Input text opinion.

y_i : Dependent variable representing justice i, where i ranges from 1 to 9 (corresponding to the nine serving justices).

$P(y_i \mid x)$: Probability of justice y_i as author, given input opinion x.

$\phi(y_i \mid x)$: Feature vector with entries corresponding to each of the n-gram features for each justice.

θ: Weight vector (the set of coefficients on each feature).

For a given document, the MaxEnt model computes a score for each justice, *i,* that is a weighted sum of the n-gram features. Using the training data, the machine learning algorithms automatically learn the parameters of the weight vector to maximize the likelihood of the training data; that is, the algorithm adjusts the weights to best "explain" the data. The form of the MaxEnt model ensures that $P(y_i \mid x)$ is between 0 and 1 and that these probability values sum to 1, meaning that the output of the model can be interpreted as a probability distribution. This relatively simple approach has been used successfully in other text classification problems.[77] We used Apache OpenNLP[78] for the MaxEnt model and WEKA,[79] two open-source machine learning software packages, for the baseline methods outlined below.

In summary, the authorship attribution system computes a probability distribution over the nine justices for a given written opinion. The justice with

---

76. *See, e.g.*, Adam L. Berger et. al, *A Maximum Entropy Approach to Natural Language Processing*, 22(1) COMPUTATIONAL LINGUISTICS 39 (1996).

77. *See, e.g.*, Adwait Ratnaparkhi, *Maximum Entropy Models for Natural Language Ambiguity Resolution*, Ph.D. thesis, University of Pennsylvania (1998).

78. "The OpenNLP library is a machine learning based toolkit for the processing of natural language text." OpenNLP is available at http://opennlp.apache.org/.

79. "[WEKA] is a collection of machine learning algorithms for data mining tasks," and is available under a GNU General Public License at http://opennlp.apache.org/.

the highest probability is used as the predicted author, as discussed in the next Part.

Experimental Evaluation: To validate our choice of the MaxEnt authorship attribution model, we compared it to other common machine learning algorithms using the same set of features. These other common machine learning algorithms are:

a. Decision trees (DT)[80]: Instead of taking a weighted sum of features, decision tree models learn deterministic rules directly from the feature set. For example, the decision tree may use the presence or absence of a particular *n*-gram to predict one justice as opposed to another. These conceptually simple models may suffer in performance because certain features might be indicative, but not determinative, of certain authors. In probabilistic models, negative evidence against a particular author can be outweighed by positive evidence in favor of the author.

b. Naïve Bayes (NB)[81] classification: By assuming that features are statistically independent, the Naïve Bayes model calculates a probability that a justice wrote an opinion by multiplying the conditional probabilities of each feature given the justice. In other words, each feature's "contribution" to the model is computed separately because it assumes that all of the features are statistically independent. As a result, the NB model is substantially simpler and faster to train than the MaxEnt model, the latter of which involves learning the weights for all of the features. However, the NB model may not perform as well because it does not attempt to search through all possible weights to maximize performance.

c. Pairwise-coupled support vector machines (SVM)[82]: Some authorship attribution applications have reported state-of-the-art results with support vector machines, which also learn weights on features using a different mathematical formulation.[83] We used pairwise-coupled SVMs, in which the opinions of each possible pair of justices are trained. The output of each classifier is a "vote" for one justice over another, and the justice with the most overall "votes" is the predicted author. One challenge related to our desiderata is that the votes may be difficult to interpret as meaningful probabilities.

Table 3 in the next Part compares the results of each of these three models with the MaxEnt model.

---

80. *See* CHRISTOPHER M. BISHOP, PATTERN RECOGNITION AND MACHINE LEARNING § 16.4 (2006).

81. *See* DANIEL JURAFSKY & JAMES H. MARTIN, SPEECH AND LANGUAGE PROCESSING: AN INTRODUCTION TO NATURAL LANGUAGE PROCESSING, COMPUTATIONAL LINGUISTICS, AND SPEECH RECOGNITION, § 20.2.2 (2009).

82. BISHOP, *supra* note 80, at ch.7.

83. For a description of the SVM implementation we used, see John C. Platt, *Fast Training of Support Vector Machines Using Sequential Minimal Optimization*, *in* ADVANCES IN KERNEL METHODS: SUPPORT VECTOR LEARNING, (Bernhard Schoelkopf et al., eds. 1998).

3. *Feature Selection*

We also decided to selectively limit which *n*-gram features were used in the authorship attribution model. Given the length and quantity of opinions, there are hundreds of thousands of possible *n*-grams in the set of Supreme Court opinions, but not all of them are likely to be useful. For instance, an *n*-gram could reflect vocabulary specific to a single case; associating it with a particular author would not be useful for predicting authorship in the validation set or in cases where the author is unknown. In addition, having too many features in our model could make it prone to "over-fitting"—the results would not generalize to opinions in our validation set.[84]

First, we considered only features that appear in a minimum of 20 opinions in our training set, thereby eliminating very case-specific or rare language. Then, we considered two feature selection methods that are commonly used for text processing: document frequency and information gain.[85] In both cases, we computed a score for each eligible *n*-gram feature and then took the highest-ranked features:

a. Document Frequency (DF) computes the frequency score of the *n*-gram simply by counting the number of documents that include the *n*-gram. We selected the 3000 most frequent unigrams, 1000 most frequent bigrams, and 1000 most frequent trigrams in the training set for our DF-based model. While DF-based feature selection is simple, it has been shown empirically to be as effective as more sophisticated methods in text classification tasks.[86] Choosing frequent *n*-grams could be a reasonable feature selection method because *n*-grams that reflect common writing styles or patterns are likely to appear in the opinions in our validation set.

b. Instead of scoring each *n*-gram feature by its frequency of appearance, Information Gain (IG) measures the contribution of a particular feature to differentiating among the justices, which is the ultimate goal of our authorship attribution model. Specifically, we compute the weighted average entropy of each feature, *f*:

$$IG(f) = P(f_{present}) \sum_{k=1}^{9} P(j_k|f_{present}) \log P(j_k|f_{present}) + P(f_{absent}) \sum_{k=1}^{9} P(j_k|f_{absent}) \log P(j_k|f_{absent})$$

where:

$P(f_{present})$: The fraction of documents that contain the *n*-gram feature.

$P(f_{absent})$: The fraction of documents that do not contain the *n*-gram

84. BISHOP, *supra* note 80, at ch. 1.

85. Yiming Yang and Jan O. Pedersen, *A Comparative Study on Feature Selection in Text Categorization*, in PROCEEDINGS OF THE 14TH INTERNATIONAL CONFERENCE ON MACHINE LEARNING (1997).

86. *Id.*

feature.

$\sum_{k=1}^{9} P(j_k|f_{present}) \log P(j_k|f_{present})$: A measure of the non-uniformity of the probability distribution of opinions authored by different justices, conditioned on the presence of the feature *f*. This definition is precisely the negative of the entropy[87] of the probability distribution—the more non-uniform the probability distribution, the higher the score.

$\sum_{k=1}^{9} P(j_k|f_{absent}) \log P(j_k|f_{absent})$: A measure of the non-uniformity of the probability distribution of opinions by justice, conditioned on the absence of feature *f*.

Once we computed these scores for every eligible *n*-gram, we chose the *n*-grams with the highest scores. For the IG model, we chose 3002 unigrams (similar to the number of unigrams in the DF model). We included bigrams and trigrams that had an IG score at least as high as one of the included unigrams. As a result, we selected 2714 bigrams and 746 trigrams into our model.

Table 2 shows examples of *n*-gram features that were selected using DF, IG, or both feature selection methods. Some of the features selected using DF, such as "the" or "at", had low IG scores because they appear in almost every document; consequently, they were not chosen as features when evaluated for information gain. Despite substantial differences in the two feature sets, it is worth noting that many of the types of *n*-grams are quite similar—they seem to encode characteristics of justices' writing styles. A reasonable hypothesis is that a model with features selected under our information gain method is more likely to yield better results, but this must be validated experimentally.

TABLE 2: EXAMPLES OF *N*-GRAM FEATURES SELECTED BY DOCUMENT FREQUENCY (DF) AND INFORMATION GAIN (IG) METHODS

Features only in DF Model (3154 features)	**Features in DF and IG models (1846 features)**	**Features only in IG model (4616 features)**
"the", "at", "exclude", "conceded", "refers to", "entitled to", "exception to the", "see no reason"	"stated,", "consequently", "declared", "assumption", "Even if", "consideration of", "and the case", "fact that the"	"Furthermore,", "troubling", "undisturbed", "evidently", "That would", "assert that", "the premise that", "is apparent that"

Experimental Evaluation: Given that both feature selection methods are commonly used in algorithmic text analysis, we present results using both DF and IG for the *Sebelius* decision. Showing the findings from both models could give greater insight into authorship in this decision.

87. *See* Claude Shannon, *Prediction and Entropy of Printed English*, 30(1) JUDICATUREBELL SYSTEMS TECHNICAL JOURNAL 50 (1951).

IV. EMPIRICAL RESULTS AND DISCUSSION

A. *Feature Sets and Classification Models*

Table 3 summarizes the performance of feature types and classification models mentioned in the previous Part. For consistency in this set of experiments, the DF method of selecting features was used for all of these models, meaning that the feature set is the same. These results were obtained by ten-fold cross-validation, in which a model is trained on 90% of the training data, the results are computed for the remaining 10%, and the 90/10 split is repeated a total of ten times to judge the performance on the entire, 451-opinion training set. The steps of cross-validation were as follows:

1. We randomly divided the set of documents into ten equal partitions.
2. We "held out" one partition and trained the authorship attribution model on the remaining nine partitions. Then, we tested the trained model on the "held out" partition.
3. We repeated step 2 once for each of the ten partitions, then combined the results. Given that there were ten partitions, this process is called ten-fold cross-validation.

It appears that the scores for correctly and incorrectly classified opinions are drawn from different, albeit overlapping, distributions for the MaxEnt-DF and MaxEnt-IG models. In general, for both models, higher output scores correspond to higher likelihood that the prediction is correct.

As seen below, the unigram features provide good prediction accuracy on their own, with some improvement when bigrams and trigrams are added as features. The maximum entropy (MaxEnt) and support vector machine (SVM) classifiers outperform the Naïve Bayes (NB) and decision tree (DT) models. Additionally, the MaxEnt model performs slightly better than the SVM and has the added property of probability distributions over the justices. Overall, these findings help justify the use of unigrams, bigrams, and trigrams as features in a MaxEnt framework in our authorship attribution model. The remaining results and analysis in this Part focus on variants of the MaxEnt model exclusively.

TABLE 3: AUTHORSHIP PREDICTION ACCURACY BY FEATURE SET AND CLASSIFIER

Features	DT	NB	SVM	MaxEnt
Unigrams	0.322	0.514	0.734	0.736
Bigrams	0.266	0.443	0.559	0.588
Trigrams	0.244	0.459	0.501	0.548
Uni/Bi/Trigrams	0.301	0.527	0.747	0.752

B. *Comparison of Feature Selection Models*

Given the strong performance and desirable properties of the MaxEnt classification model, we evaluated the model on our 117-opinion validation set

using the document frequency (MaxEnt-DF) and information gain (MaxEnt-IG) feature selection methods. We obtained the features and model weights from the 451-opinion training set. Table 4 compares the performance of the MaxEnt model using the document frequency (MaxEnt-DF) and information gain (MaxEnt-IG) methods of feature selection. IG results in a performance of 81.2% (95 out of 117 cases correct), while DF has a lower accuracy at 76.1% (89 out of 117). The somewhat lower performance of MaxEnt-DF is likely because some features do not help to differentiate the justices and may merely add noise to the model.

TABLE 4: PERFORMANCE OF MAXENT MODELS WITH DOCUMENT FREQUENCY (DF) AND INFORMATION GAIN (IG) FEATURE SELECTION

	MaxEnt-DF	**MaxEnt-IG**
Accuracy on Test Set	0.761 (89/117)	0.812 (95/117)

C. *Interpreting Authorship Attribution Model Scores*

In addition to predicting a justice, the MaxEnt models also provide an output score for each justice that can be interpreted as a probability. Figure 1 shows the distribution of maximum scores for correctly and incorrectly classified opinions in the training set. In order to characterize the behavior of the model scores on our entire set of opinions, we obtained these results through the machine learning technique of ten-fold cross-validation, similar to the approach used to compare the different learning models.

Instead of simply taking the justice with the highest probability as the prediction, a more refined prediction system that "abstains" (makes no prediction) below a certain output probability can be constructed. The ratio of correctly predicted cases to incorrectly predicted cases increases as this threshold increases at the expense of a larger number of abstentions. This result is visualized in Figure 2 and provides another illustration of the differences between the MaxEnt-DF and MaxEnt-IG models. For example, in the MaxEnt-IG model, a threshold probability of 0.43 results in 90% prediction accuracy, with abstentions on just 19.3% of all cases. In contrast, to achieve 90% prediction accuracy, the MaxEnt-DF model must set a threshold of 0.52, abstaining on 35.1% of all cases. Overall, these plots illustrate the probabilistic nature of our authorship attribution model. Based on the desired application of the model, one could set different abstaining thresholds, depending on the level of confidence desired.

FIGURE 1: HISTOGRAM OF PROBABILITIES OF MOST PROBABLE JUSTICE FOR MAXENT-DF (LEFT) AND MAXENT-IG (RIGHT) MODELS.

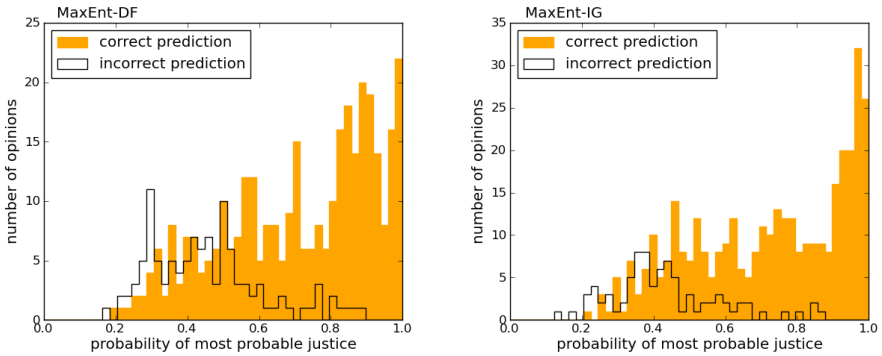
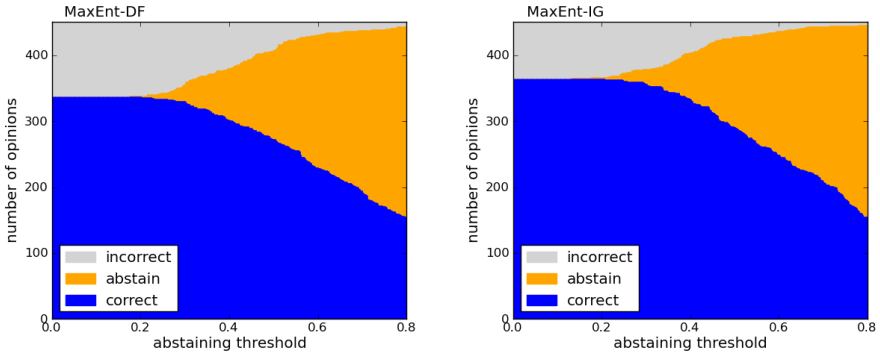


FIGURE 2: EFFECT OF ABSTAINING THRESHOLD ON SIZE OF CORRECT, INCORRECT, AND ABSTAINING CLASSES OF OPINIONS FOR MAXENT-DF (LEFT) AND MAXENT-IG MODELS (RIGHT).



D. *Insights on Writing Styles*

To provide some insight into how our authorship attribution system predicts which justice wrote an opinion, Table 5 shows some of the most predictive unigrams, bigrams, and trigrams for each justice from the MaxEnt-IG model. This table was computed by determining *n*-grams that appear disproportionately more often for each justice. Some noteworthy insights include:

1. The highly predictive *n*-grams are largely topic-invariant function terms; that is, the informative features more frequently reflect the writing style of the justice as opposed to specific subjects. For example, the term "consequently" appears in 99 different opinions in our training dataset; Justice Breyer wrote 79 of these opinions.
2. Some predictive *n*-grams begin with capitalized words, including "For one thing" (Breyer), "Notably," (Ginsburg), and "The question is" (Kennedy). These correspond to words at the beginning of sentences, indicating that how different justices start sentences provides clues about authorship.
3. Some predictive *n*-grams include punctuation like commas and periods, including "reason stated, the" (Ginsburg), "the first place." (Roberts), and "foregoing reasons," (Thomas). By not eliminating punctuation from the text, the authorship attribution model is able to leverage these stylistic features.

TABLE 5: INFORMATIVE FEATURES BY JUSTICE

	Highly predictive *n*-grams		
Justice	**Unigrams**	**Bigrams**	**Trigrams**
Alito	"fundamentally", "widely", "regarded"	"set out", "noted above", "is generally"	"set out in", "and we have", "the decision of"
Breyer	"consequently", "Hence", "thing,"	"can find", "wrote that", "For one"	"in respect to", "For one thing", "That is because"
Ginsburg	"Notably,", "observed,", "stated,"	"reasons stated,", "stated, the", "case concerns"	"stated, the judgment", "reasons stated, the"
Kagan	"enables", "earlier," "matters."	"result is", "after all,", "the theory"	"do not think", "Court has never", "even when the"
Kennedy	"however.", "responsibilities", "Though"	"It held", "so the", "or she"	"The question is", "as a general", "he or she"
Roberts	"pertinent", "accordingly", "Here"	"first place.", "only be", "given that"	"the first place.", "without regard to", "a general matter,"
Scalia	"utterly", "thinks", "finally"	"Of course", "since it", "is entirely"	"That is not", "the present case", "is hard to"
Sotomayor	"observes", "lawsuits", "heightened"	"Committee on", "federal and", "correct that"	"circumstances in which", "see also, Brief", "federal and state"
Thomas	"Therefore," "However,", "explaining"	"address whether", "foregoing reasons," "Court holds"	"hold that it", "For the foregoing", "the foregoing reasons"

E. *Controlling for Clerks*

The training/test split in the experiments above was random with respect to the year in which the opinion was written, meaning that, at least to some extent, a justice's writing style in a given year can be predicted from his or her writings in other years. Our model does not explicitly consider the role of law clerks, who typically serve year-long terms, in the writing process; rather, it assumes that the features of a justice's writing are similar from year to year.

To test this assumption, we once again applied the cross-validation technique (similar to how we studied the model confidence scores) and divided the set of documents in two different ways. Specifically, we took signed

opinions from the Roberts Court in the 2005-2006, 2006-2007, 2007-2008, 2008-2009, 2009-2010, and 2010-2011 sessions and compared the performance of our model in two ways:

1. For each of the six annual sessions, we trained the MaxEnt-IG model on the other five sessions and validated on cases from the omitted session.
2. The signed opinions were randomly divided into six partitions, irrespective of the year. For each partition, we trained the MaxEnt-IG model on the other five partitions and validated on opinions from the omitted partition.

In both cases, we aggregated the results from each of the six runs, as shown in Table 6. Evidently, training on other years has only a slight adverse impact on the accuracy of the model. This may suggest that a different set of clerks have some impact on a justice's writing style, although other factors, such as a justice's own drift in writing style, may contribute to this result. Overall, though, the higher performance of the randomized model suggests that the combination of training data from other years and the same year works well for predicting authorship.

TABLE 6: PREDICTION ACCURACY OF MODELS TRAINED ON OPINIONS FROM DIFFERENT YEARS

Partition	**Accuracy**
Year-based	74.8% (306/412)
Random	78.9% (325/412)

F. *Authorship Prediction for* Sebelius

To infer authorship in the *Sebelius* decision, we trained the MaxEnt-DF and MaxEnt-IG models on the 568 cases in our dataset and ran them on the majority opinion signed by Chief Justice Roberts and the joint dissent. Figure 3 and Figure 4 illustrate the resulting probability distributions of the two models. Both MaxEnt-DF and MaxEnt-IG strongly predict Chief Justice Roberts for the majority opinion. For the joint dissent, the MaxEnt-DF model states that Justice Kennedy is the predicted author, but the probability distribution is not as peaked—Justice Scalia has the second-highest probability. Meanwhile, the MaxEnt-IG model is much more confident in Justice Scalia as the author of the joint dissent. Overall, these findings are sensible: Chief Justice Roberts signed the majority opinion, while Kennedy and Scalia are listed as authors of the joint dissent. Both models support the hypothesis that Kennedy and Scalia were authors and prime actors in writing the dissent, and refute the hypothesis that Roberts authored both opinions. The output of the model in the *Sebelius* decision is an example of the non-partisan, quantitative analysis that the authorship attribution system can provide.

FIGURE 3: AUTHORSHIP ATTRIBUTION MODEL PREDICTION FOR SEBELIUS MAJORITY OPINION BY MAXENT-DF (LEFT) AND MAXENT-IG (RIGHT).

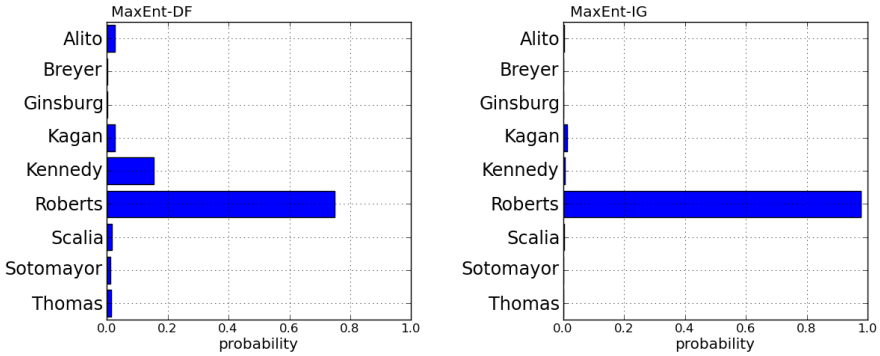
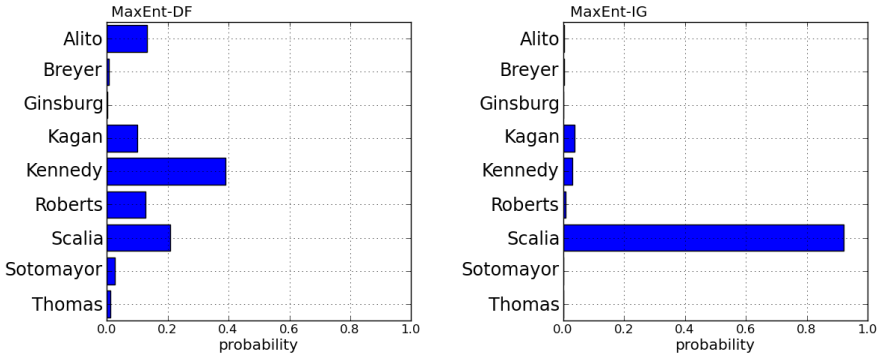


FIGURE 4: AUTHORSHIP ATTRIBUTION MODEL PREDICTION FOR SEBELIUS JOINT DISSENT BY MAXENT-DF (LEFT) AND MAXENT-IG (RIGHT).



G. *Comparison to Predictions by Domain Experts*

To gauge the performance of our authorship attribution system to expert human judgment, we received input from 11 individuals close to the Supreme Court (as past law clerks or lawyers who have argued at least one case before the Supreme Court).[88] We asked each of the respondents to provide their best, informed guess of the authorship of the joint dissent, annotated, if possible, by section. Out of the 11 responses, there were nine who concluded that Kennedy participated in some way, nine for Roberts, six for Scalia, and three for Alito; no other justices were mentioned. The predictions of our model on this case generally seem to be in agreement with these domain experts, although it is worth noting that the MaxEnt-IG model is more confident in Scalia than the domain experts. A summary of the responses from each of the respondents is listed in Table 7.

88. All of the respondents asked to remain anonymous.

TABLE 7: PREDICTIONS OF AUTHORSHIP OF DISSENTING OPINION BY DOMAIN EXPERTS

Respondent	Predicted authors (ranked in order of level of contribution)
1	Kennedy, Roberts
2	Scalia, Roberts
3	Roberts, Kennedy
4	Roberts, Scalia, Kennedy, and Alito
5	Kennedy, Alito
6	Scalia, Kennedy, Roberts
7	Scalia, working loosely from a draft by Roberts
8	Kennedy, working loosely from a draft by Roberts
9	Roberts, Kennedy, Scalia
10	Kennedy, Alito, Roberts
11	Scalia, Kennedy

H. *Section-by-Section Analysis*

Given that some respondents provided predictions by section, we tested our MaxEnt models on the joint dissent divided into nine sections. We emphasize that the models were not trained on portions of opinions; however, understanding the contributions of different justices to the constituent parts of an opinion could be useful. Sections are frequently delineated according to a precise legal issue and, theoretically, one justice could contribute his or her treatment of a specific legal issue to be incorporated into an opinion written by another justice.

The top predictions and predictions for the MaxEnt-DF and MaxEnt-IG models for each of *Sebelius*'s sections are shown in Table 8. The results are somewhat noisy—our respondents did not predict Breyer or Thomas as authors of the joint dissent, and the two models did not agree on every section. However, consistent with its prediction for the entire opinion, the MaxEnt-IG model predicted Scalia for most of the sections. Additionally, none of the predictions suggest that Roberts is the top author, which may lend further evidence against the theory that Roberts authored the dissenting opinion.

TABLE 8: PREDICTION OF AUTHORSHIP OF DISSENTING OPINION BY SECTION

	Predicted author	
Section	**MaxEnt-DF**	**MaxEnt-IG**
Introduction	Scalia (0.841)	Scalia (0.544)
Sec. 1 Introduction	Scalia (0.365)	Scalia (0.235)
Sec. 1A	Breyer (0.406)	Scalia (0.284)
Sec. 1B	Scalia (0.730)	Scalia (0.613)
Sec. 1C	Kennedy (0.904)	Scalia (0.590)
Sec. 2	Scalia (0.552)	Scalia (0.891)
Sec. 3	Thomas (0.541)	Scalia (0.344)
Sec. 4	Scalia (0.770)	Alito (0.283)
Sec. 5	Alito (0.385)	Kennedy (0.738)

V. AUTHORSHIP PREDICTIONS FOR PER CURIAM OPINIONS OF THE ROBERTS COURT

Finally, we tested the MaxEnt-IG model on 65 per curiam opinions of the Roberts Court since 2005, with the goal of inferring the authorship of these unsigned opinions. For each opinion, we trained a MaxEnt model using data from the nine sitting justices at the time; for example, in 2006, the training set consists of opinions from Justices Stevens, Scalia, Kennedy, Souter, Thomas, Ginsburg, Breyer, and Alito, along with Chief Justice Roberts. It is worth noting that the model's output probabilities for the most probable justice is often fairly low; for example, if we supplied a cutoff threshold of 0.43 to have 90% confidence in our prediction (as per Figure 2), the model would choose to "abstain" on predicting a justice in many of these cases.

TABLE 9: PREDICTED AUTHORSHIP OF ROBERTS COURT PER CURIAM DECISIONS

Date of Decision	**Case**	**Highest probability justice**	**Second-highest probability justice**	**Third-highest probability justice**
10/5/2005	*Dye v. Hofbauer*	Kennedy (0.466)	Scalia (0.165)	Ginsburg (0.093)
10/17/2005	*Schiro v. Smith*	O'Connor (0.207)	Thomas (0.192)	Scalia (0.152)
10/31/2005	*Eberhart v. United States*	Thomas (0.268)	Scalia (0.189)	Ginsburg (0.165)
10/31/2005	*Kane v. Garcia Espitia*	Scalia (0.190)	Thomas (0.174)	O'Connor (0.145)
11/28/2005	*Bradshaw v. Richey*	O'Connor (0.325)	Scalia (0.294)	Thomas (0.115)
1/23/2006	*Wisconsin Right to Life, Inc. v. Federal Election Commission*	Roberts (0.210)	O'Connor (0.151)	Stevens (0.129)
2/21/2006	*Ash v. Tyson Foods, Inc.*	Kennedy (0.326)	Scalia (0.304)	Thomas (0.149)
2/21/2006	*Lance v. Dennis*	Scalia (0.351)	Ginsburg (0.146)	Thomas (0.122)
2/21/2006	*Ministry of Defense and Support v. Elahi*	Breyer (0.397)	Scalia (0.209)	Thomas (0.179)
4/17/2006	*Gonzales v. Thomas*	Breyer (0.715)	Thomas (0.081)	Scalia (0.076)
4/24/2006	*Salinas v. United States*	Roberts (0.171)	Breyer (0.152)	Scalia (0.151)
6/5/2006	*Whitman v. Department of Transportation*	Kennedy (0.248)	Scalia (0.180)	Souter (0.135)

6/19/2006	*Youngblood v. West Virginia*	Scalia (0.193)	Ginsburg (0.182)	Thomas (0.126)
10/20/2006	*Purcell v. Gonzalez*	Kennedy (0.574)	Ginsburg (0.142)	Stevens (0.058)
1/9/2007	*Burton v. Stewart*	Thomas (0.452)	Roberts (0.229)	Scalia (0.182)
3/5/2007	*Lance v. Coffman*	Scalia (0.306)	Roberts (0.300)	Alito (0.081)
5/21/2007	*Los Angeles County v. Rettele*	Scalia (0.301)	Kennedy (0.189)	Stevens (0.111)
5/21/2007	*Roper v. Weaver*	Thomas (0.330)	Kennedy (0.127)	Ginsburg (0.117)
6/4/2007	*Erickson v. Pardus*	Ginsburg (0.335)	Scalia (0.214)	Thomas (0.146)
11/5/2007	*Allen v. Siebert*	Thomas (0.292)	Scalia (0.261)	Roberts (0.153)
1/7/2008	*Arave v. Hoffman*	Ginsburg (0.212)	Thomas (0.154)	Kennedy (0.152)
1/7/2008	*Wright v. Van Patten*	Thomas (0.517)	Scalia (0.135)	Ginsburg (0.068)
8/5/2008	*Medellín v. Texas*	Kennedy (0.186)	Breyer (0.164)	Scalia (0.140)
10/14/2008	*Moore v. United States*	Thomas (0.203)	Scalia (0.167)	Breyer (0.132)
12/2/2008	*Brunner v. Ohio Republican Party*	Ginsburg (0.153)	Kennedy (0.139)	Scalia (0.126)
12/2/2008	*Hedgpeth v. Pulido*	Thomas (0.256)	Roberts (0.211)	Breyer (0.140)
1/21/2009	*Spears v. United States*	Scalia (0.531)	Roberts (0.201)	Thomas (0.121)
1/26/2009	*Nelson v. United States*	Thomas (0.209)	Scalia (0.193)	Roberts (0.138)
6/1/2009	*CSX Transportation, Inc. v. Hensley*	Roberts (0.179)	Kennedy (0.157)	Scalia (0.143)
6/9/2009	*Indiana State Police Pension Trust v. Chrysler LLC*	Roberts (0.186)	Ginsburg (0.163)	Alito (0.121)
10/20/2009	*Concoran v. Levenhagen*	Scalia (0.210)	Breyer (0.148)	Kennedy (0.140)
11/9/2009	*Bobby v. Van Hook*	Breyer (0.206)	Kennedy (0.203)	Ginsburg (0.197)
11/16/2009	*Wong v. Belmontes*	Kennedy (0.428)	Scalia (0.229)	Roberts (0.137)
11/30/2009	*Porter v. McCollum*	Thomas (0.229)	Scalia (0.191)	Kennedy (0.169)

12/7/2009	*Michigan v. Fisher*	Roberts (0.244)	Scalia (0.178)	Alito (0.122)
1/11/2010	*McDaniel v. Brown*	Thomas (0.768)	Kennedy (0.074)	Scalia (0.038)
1/13/2010	*Hollingsworth v. Perry*	Kennedy (0.787)	Scalia (0.066)	Stevens (0.040)
1/19/2010	*Presley v. Georgia*	Kennedy (0.619)	Stevens (0.089)	Scalia (0.068)
1/19/2010	*Wellons v. Hall*	Scalia (0.440)	Thomas (0.276)	Ginsburg (0.064)
2/22/2010	*Thaler v. Haynes*	Alito (0.248)	Thomas (0.238)	Ginsburg (0.179)
2/22/2010	*Wilkins v. Gaddy*	Scalia (0.243)	Ginsburg (0.233)	Thomas (0.212)
3/1/2010	*Kiyemba v. Obama*	Roberts (0.181)	Scalia (0.161)	Ginsburg (0.154)
5/24/2010	*Jefferson v. Upton*	Scalia (0.330)	Breyer (0.239)	Thomas (0.231)
6/7/2010	*United States v. Juvenile Male*	Thomas (0.249)	Roberts (0.194)	Scalia (0.141)
6/29/2010	*Sears v. Upton*	Thomas (0.223)	Breyer (0.159)	Scalia (0.157)
11/8/2010	*Wilson v. Corcoran*	Scalia (0.324)	Thomas (0.201)	Ginsburg (0.186)
1/10/2011	*Madison County v. Oneida Indian Nation*	Thomas (0.190)	Kennedy (0.139)	Ginsburg (0.139)
1/24/2011	*Swarthout v. Cooke*	Scalia (0.254)	Roberts (0.190)	Thomas (0.184)
3/21/2011	*Felkner v. Jackson*	Thomas (0.359)	Roberts (0.211)	Kennedy (0.177)
5/2/2011	*Bobby v. Mitts*	Thomas (0.528)	Scalia (0.124)	Roberts (0.123)
7/7/2011	*Leal Garcia v. Tecas*	Scalia (0.240)	Roberts (0.234)	Breyer (0.138)
10/31/2011	*Cavazos v. Smith*	Kennedy (0.334)	Ginsburg (0.195)	Scalia (0.183)
11/7/2011	*Bobby v. Dixon*	Scalia (0.373)	Kennedy (0.213)	Ginsburg (0.139)
11/7/2011	*KPMG LLP v. Cocchi*	Thomas (0.285)	Kennedy (0.254)	Scalia (0.111)
12/12/2011	*Hardy v. Cross*	Thomas (0.247)	Alito (0.223)	Ginsburg (0.186)
1/20/2012	*Perry v. Perez*	Kennedy (0.344)	Roberts (0.230)	Scalia (0.184)
1/23/2012	*Ryburn v. Huff*	Alito (0.212)	Ginsburg (0.185)	Scalia (0.176)
2/21/2012	*Wetzel v. Lambert*	Roberts (0.239)	Thomas (0.190)	Kennedy (0.177)

5/29/2012	*Coleman v. Johnson*	Thomas (0.225)	Kennedy (0.203)	Breyer (0.164)
5/29/2012	*Marmet Health Care Center, Inc. v. Brown*	Thomas (0.324)	Kennedy (0.195)	Alito (0.139)
5/29/2012	*Parker v. Matthews*	Kennedy (0.265)	Ginsburg (0.181)	Scalia (0.134)
6/25/2012	*American Tradition Partnership, Inc. v. Bullock*	Thomas (0.232)	Roberts (0.161)	Kennedy (0.147)
9/25/2012	*Tennant v. Jefferson County*	Roberts (0.603)	Kennedy (0.119)	Ginsburg (0.076)
11/5/2012	*Lefemine v. Wideman*	Ginsburg (0.267)	Roberts (0.188)	Thomas (0.182)
11/26/2012	*Nitro-Lift Technologies, LLC v. Howard*	Thomas (0.472)	Scalia (0.248)	Kennedy (0.080)

Assuming these predictions are accurate, they are provocative. Justices commonly described as "conservative" are predicted authors of 45 out of the 65 per curiam opinions (69.2%). Justices commonly described as "conservative-swing" are predicted authors of 13 of the remaining 20 opinions—11 for Justice Kennedy and 2 for Justice O'Connor. Thus, conservative or conservative-swing justices are predicted authors of 58 out of the 65 per curiam opinions (89.2%). Justices commonly described as liberal are predicted authors of only 7 of the opinions (10.8%). Table 10 provides a yearly break down of the predicted authors by their ideology.

TABLE 10: PREDICTED AUTHOR IDEOLOGY OF PER CURIAM OPINIONS BY YEAR

Year	**Total**	**Conservative**	**Conservative-swing**	**Liberal**
2012	2	1 (50%)	0	1 (50%)
2011	12	9 (75%)	3 (25%)	0
2010	14	10 (71.4%)	3 (21.4%)	1 (7.1%)
2009	14	10 (71.4%)	3 (21.4%)	1 (7.1%)
2008	7	6 (85.7%)	0	1 (14.3%)
2007	4	2 (50%)	1 (25%)	1 (25%)
2006	5	3 (60%)	1 (20%)	1 (20%)
2005	14	7 (50%)	5 (35.7%)	2 (14.3%)

Excluding 2012, in which there were only two per curiam opinions, conservative and conservative-swing justices in the Roberts Court are predicted authors of between 75% and 100% of the per curiam decisions per year.

CONCLUSION

Machine learning techniques can be used to attribute authorship of judicial opinions. We have demonstrated that word-level features can distinguish authorship with substantial accuracy. The inferred authorship for the opinions in *Sebelius* provides unambiguous quantitative support for one theory of authorship offered in the media. Applying these methods to the unsigned per curiam opinions of the Roberts Court yields provocative results. The performance of the model on test opinions, along with the stylistic features that it uses to determine performance, suggests that it could be useful for other courts. Overall, our work underscores the broad applicability of natural language processing tools to yield quantitative insights into issues traditionally studied only qualitatively and manually.