ESSAY

THE SUPREME COURT FORECASTING PROJECT: LEGAL AND POLITICAL SCIENCE APPROACHES TO PREDICTING SUPREME COURT DECISIONMAKING

*Theodore W. Ruger, Pauline T. Kim, Andrew D. Martin, & Kevin M. Quinn**

This Essay reports the results of an interdisciplinary project comparing political science and legal approaches to forecasting Supreme Court decisions. For every argued case during the 2002 Term, we obtained predictions of the outcome prior to oral argument using two methods—one a statistical model that relies on general case characteristics, and the other a set of independent predictions by legal specialists. The basic result is that the statistical model did better than the legal experts in forecasting the outcomes of the Term's cases: The model predicted 75% of the Court's affirm/reverse results correctly, while the experts collectively got 59.1% right. These results are notable, given that the statistical model disregards information about the specific law or facts of the cases. The model's relative success was due in large part to its ability to predict more accurately the important votes of the moderate Justices (Kennedy and O'Connor) at the center of the current Court. The legal experts, by contrast, did best at predicting the votes of the more ideologically extreme Justices, but had difficulty predicting the centrist Justices. The relative success of the two methods also varied by issue area, with the statistical model doing particularly well in forecasting "economic activity" cases, while the experts did comparatively better in the "judicial power" cases. In addition to reporting the results in detail, the Essay explains the differing methods

---

* Theodore W. Ruger is Associate Professor of Law and Pauline T. Kim is Professor of Law at Washington University in St. Louis. Andrew D. Martin is Assistant Professor in the Department of Political Science at Washington University. Kevin M. Quinn is Assistant Professor in the Harvard University Department of Government. This study has benefited from helpful comments from Theodore Eisenberg, Lee Epstein, Tracey George, Mitu Gulati, Nancy Staudt, Mark Tushnet, and participants at the Colloquium on Law, Economics, and Politics at New York University School of Law, at law faculty workshops at Boston University Law School, Fordham Law School, the University of Pennsylvania Law School, Washington University Law School, and at the Workshop on Empirical Research in the Law at Washington University. We thank Michael Cherba, Nancy Cummings, Alison Garvey, Nick Hershman, Winston Calvert, Robyn Rimmer, and Sahmon Torabi for their research and administrative assistance. We owe a special debt of gratitude to the eighty-three legal experts who generously agreed—on an entirely volunteer basis—to spend time participating in this experiment during the past year. See Appendix B for a list of experts. This project is supported in part by National Science Foundation grants SES 01-35855 and SES 01-36679. The Foundation bears no responsibility for the results or conclusions. All calculations are by the authors, based on underlying data on file with the authors and the *Columbia Law Review*. A condensed and peer-reviewed version of this study is the subject of a symposium forthcoming in *Perspectives on Politics*.

of prediction used and explores the implications of the findings for assessing and understanding Supreme Court decisionmaking.

INTRODUCTION

"Our business is prophecy, and if prophecy were certain, there would not be much credit in prophesying."[1]

The 2002 Term of the Supreme Court underscored two essential, and fairly obvious, features of the institution and its place in American political society. The Court is often important, and it is occasionally surprising. The Court's decisions impact a diverse array of vital economic, social, and structural questions. To mention just a few of the Term's cases, the Court declared rules about the constitutionality of affirmative action,[2] the right to engage in consensual homosexual sodomy,[3] various free speech rights,[4] and the contours of the federal-state allocation of authority.[5] Furthermore, the Court's decisions in these and other areas are frequently hard to predict in advance, at least in the eyes of many lawyers, legal academics, and specialized journalists who follow the Court closely. Commentary on the 2002 Term has described it as "stunn[ing],"[6] "a [s]urprise,"[7] "startling,"[8] "idiosyncratic,"[9] "counterintuitive," and as "upending the expectations of those who watch and analyze it."[10]

Our study joins this discussion of the Supreme Court and its 2002 Term, but from a different temporal perspective than most legal and political science commentary on the Court. Rather than focus retrospectively, and proceed to analyze, critique, quantify, regress, debunk, reconcile, classify, or applaud some set of the Court's past decisions, we instead applied two different methods to predict the outcome of every case argued in the Term. In advance of the oral argument date, we obtained predicted outcomes using two methods—one a statistical model that forecasts outcomes based on six general case characteristics, and the other a set of independent predictions from a large group of legal specialists, each making particularized assessments of one or more cases. We discuss these methods and the results, as well as the study's implications and limitations, at length later in this Essay, but the condensed version is that,

---

1. Max Radin, The Theory of Judicial Decision: Or How Judges Think, 11 A.B.A. J. 357, 362 (1925).

2. Grutter v. Bollinger, 123 S. Ct. 2325 (2003).

3. Lawrence v. Texas, 123 S. Ct. 2472 (2003).

4. United States v. Am. Library Ass'n, 123 S. Ct. 2297 (2003).

5. Nev. Dep't of Human Res. v. Hibbs, 123 S. Ct. 1972 (2003).

6. Charles Lane, Civil Liberties Were Term's Big Winner: Supreme Court's Moderate Rulings a Surprise, Wash. Post, June 29, 2003, at A1.

7. Id.

8. Linda Greenhouse, In a Momentous Term, Justices Remake the Law, and the Court, N.Y. Times, July 1, 2003, at A1.

9. Tony Mauro, It's a Mad, Mad, Mad, Mad Court: Justices Upended Expectations in 2002–2003 Term, Tex. Law., July 7, 2003, at 12.

10. Id.

somewhat to our surprise,[11] the machine did significantly better at predicting outcomes than did the experts. While the experts correctly forecast outcomes in 59.1% of cases, the machine got a full 75% right.

The prospective orientation of this study is unusual—but the comparative study of Supreme Court decisionmaking by legal and political science scholars is not. The body of work on the Supreme Court in both disciplines is large and diverse, and taken together embraces a wide range of motivational theories about how, and why, the Justices decide cases as they do.[12] Some of these accounts explore the potential constraints on judicial discretion supplied by case law, text, and history, others focus on broader interpretive theories, others highlight the Justices' individual policy preferences or social backgrounds, and others regard the Court and its Justices as operating strategically in a complex institutional setting that can influence outcomes. Most of these positions have adherents in both the law and political science academies, and many scholars in both disciplines regard several, if not all, of the aforementioned factors as important influences on judicial decisionmaking. Although legal academics as a group place relatively more weight on doctrine, text, and legal principle in their analysis of judicial behavior, and political scientists tend to stress attitudinal and institutional explanations more heavily, both disciplines are highly internally heterogeneous in terms of motivational theory.

Much plainer than this theoretical picture are clear differences in the methods that legal academics and political scientists typically use to study the Court. The basic distinctions are several. The first relates to the component of the Court's output that is the focal point of study. Most legal academics direct significant attention to the internal content of the Court's opinions in a given area. This generality applies not just to those who would justify or reconcile particular doctrinal or historical statements by the Court, but also to doctrine skeptics and critics who often seek to undermine the Court's rationales by exposing flaws in the expressed judicial reasoning through close analysis and critique. Conversely, political scientists have tended to focus more heavily (and often exclusively) on the Court's basic results ("affirm" or "reverse") and the Justices' individual votes in support of or dissent from such outcomes. Harold Spaeth, long a leading proponent of the attitudinal model of judicial decisionmaking in the political science academy, expressed this distinction sharply in a debate with a more doctrine-focused colleague decades ago: "I find the key to judicial behavior in what the justices do, Professor Mendelson in what they say. I focus upon their votes, he upon

11. And perhaps chagrin, at least for the two of us who claim some legal expertise ourselves.

12. We summarize a fraction of this literature infra Part I. Not treated here is the large corpus of normative scholarship in law and (to a lesser degree) in political science about how Justices *should* go about deciding cases.

their opinions."[13] Many political scientists who dispute Spaeth's attitudinal conclusions nonetheless share his initial approach to assessing Court decisionmaking by looking first at voting results.[14]

Another general difference in method is the number of cases that are subject to a given effort of analytical synthesis. Many legal scholars who seek to understand the Court study a handful of cases in a particular doctrinal area, and weight the cases unevenly, placing analytical primacy on the "leading" holdings.[15] Such focus is driven by the prevailing conventions of legal scholarship. Close reading and analysis of opinion content takes time, and convincing explanation or refutation even longer, placing practical limits on the number of holdings a legal scholar can meaningfully synthesize for analytical purposes. The subspecialization of the legal academy also leads to a narrower focus.[16] A very different baseline method exists in political science study of the Court. It is commonplace for a quantitative political science study to take account of several dozen or even several hundred cases. And in most cumulative studies, no case is given extra weight as a "leading" case; instead all are weighted equally for analytical purposes. Moreover, to the extent political scientists look at subject matter, it is often in more general categories, like "economic regulation," or "civil liberties," rather than the narrow doctrinal categories, such as "ERISA law" or "search and seizure law," that occupy legal scholars.

For all of this methodological and theoretical disagreement, however, virtually all legal and political science scholarship on the Supreme Court is retrospective in nature.[17] Whether analyzing a single case, a single Term, an entire area of doctrine, or even every Court decision over

---

13. Harold J. Spaeth, Jurimetrics and Professor Mendelson: A Troubled Relationship, 27 J. Pol. 875, 879 (1965).

14. Even those neoinstitutional political science scholars who do look within judicial opinions often treat their content as evidence of specific strategic choices made by the Justices. See, e.g., Forrest Maltzman, James F. Spriggs & Paul J. Wahlbeck, Strategy and Judicial Choice: New Institutionalist Approaches to Supreme Court Decision-Making, *in* Supreme Court Decision-Making: New Institutionalist Approaches 43, 47 (Cornell W. Clayton & Howard Gillman eds., 1999) (discussing impact of Justices' strategic interaction on opinion content).

15. The history of the *Harvard Law Review*'s annual Foreword on the previous Court Term exemplifies this feature of legal scholarship about the Supreme Court. Only one Foreword in thirty-seven years has ever mentioned, much less analyzed, every case in the preceding Term. See William N. Eskridge, Jr. & Philip P. Frickey, The Supreme Court, 1993 Term—Foreword: Law as Equilibrium, 108 Harv. L. Rev. 26 (1994). Likewise, the *Supreme Court Review*, a peer-reviewed journal of legal scholarship on the Court, typically features analytical pieces centered on particular holdings and doctrines.

16. A constitutional scholar would probably not examine an ERISA case, an ERISA scholar might not take account of a FERC case, and a FERC expert might ignore a habeas case decided contemporaneously.

17. This is obviously true of most legal critiques of particular decisions or sets of decisions, but it is also true even of political science models that make claims of "prediction." These models, discussed infra Part I, typically regress past data sets to assess consistency with various motivational hypotheses, and although they speak in terms of

several decades, those who study the Court typically apply competing explanatory frameworks to a set of existing historical facts, namely the Court's results, or opinions, or both. This is neither surprising nor inappropriate, but neither is it necessarily intrinsic in the study of a multifactorial phenomenon like Supreme Court decisionmaking.[18] What is notable, in light of all the attention focused on the Court, is that few have tried to systematically predict its decisions prospectively. Given the high economic, social, and political importance of the Court's decisions, a model that could prospectively forecast decisionmaking at a high rate of accuracy would be an invaluable tool to litigants and Court-watchers, even if the model itself were incompletely theorized. But prediction also has the potential to advance explanation by verifying, undermining, or modifying preexisting conceptions of the best ways to study the Court and understand how the Justices arrive at their decisions.

Our study compares two distinct methods of forecasting Supreme Court action, each drawing on the insights and strengths of a different discipline. Thus, the two prediction methods diverge dramatically in terms of methodology, and in this sense embody many of the differences between law and political science discussed above. The most notable distinction inheres in the level of generality the two methods employ. The statistical model looks at only a handful of case characteristics, each of them gross features easily observable without specialized legal expertise, and builds on general patterns ascertained from all 628 cases decided by the Rehnquist Court since 1994 and prior to the 2002 Term. The model is indifferent to many of the specific legal and factual aspects of the cases, instead predicting outcomes based on the same six (and only six) observable characteristics of each case.[19] The legal experts, by contrast, utilized particularized knowledge, such as the specific facts of the case or statements by individual Justices in similar cases. We did not constrain the experts to consider only "legal" factors that might drive the Court's decision. But although many considered nonlegal factors such as the Justices' policy preferences, the experts, unlike the statistical model, *could* (and did) consider particular case law and specific constitutional or statutory texts and were thus able to particularize their analysis with regard to single cases in a way that the model was not.

The basic result of our study is that the statistical model did better by a fair margin in forecasting the outcomes of last Term's cases: The

---

"predictive" accuracy, what they do is more technically called "postdiction." There have been, however, a few more overt prediction efforts that we note in the next section.

18. Other disciplines that combine theory and retrospective empirical observation, such as economics or medicine, also incorporate forecasting experiments that provide some additional evidence in support or refutation of general explanatory theories.

19. The case variables are: (1) circuit of origin; (2) issue area of the case; (3) type of petitioner (e.g., the United States, an employer, etc.); (4) type of respondent; (5) ideological direction (liberal or conservative) of the lower court ruling; and (6) whether the petitioner argued that a law or practice is unconstitutional. See infra Part II.B for a more detailed description of the model.

model predicted 75% of the Court's affirm/reverse results correctly, while the experts collectively got 59.1% right. Part III below examines this result and more specific findings of interest including the model's notable relative success at predicting the important votes of the moderate Justices (Kennedy and O'Connor) at the center of the current Court and its high success rate in certain general issue areas. Two earlier sections elaborate on the study's motivation (Part I) and methodological design (Part II).

This experiment captures only one specific Term and only one specific group of Justices, cases, and experts. The results might well be different in a different Term or with different experts. But for the 2002 Term, the model achieved notable success by utilizing a set of factors that appear to correlate with the Justices' decisionmaking. That a forecasting machine that is indifferent to specific doctrine and text can predict cases so well is interesting, surprising, and worthy of further thought. Moreover, as discussed in Part IV, the statistical model is in some sense based on spatial voting models, and as such, is consistent with decades of work in political science. Despite significant skepticism about the constraining effects of doctrine and text at the Supreme Court level, law professors still tend to think about individual Supreme Court cases in relatively particularistic legal terms. The model's success at using a much more general set of case factors to predict outcomes offers insights for all those who study and practice before the Court. We discuss these implications in Part IV.

I. Historical and Theoretical Background

Just over a hundred years ago, Holmes announced his "prediction" theory of law, explaining that "[t]he prophesies of what the courts will do in fact, and nothing more pretentious, are what I mean by the law."[20] This formulation remains highly contested in several ways, but in one particular sense—as a theoretical response to the classical legal thought of the late nineteenth century—its impact was pervasive. Holmes and his followers undermined the classical notion of law as a set of static, natural, and apolitical rules that could be mechanically discerned and applied by judges,[21] and in so doing helped to change the way in which American scholars regard the law and the legal process. Much of law is, in the modern conception, something that political society makes, and judges play some part in the making.

---

20. Oliver Wendell Holmes, Jr., The Path of the Law, 10 Harv. L. Rev. 457, 461 (1897). Elsewhere in his address, Holmes reiterated the point: "The object of our study, then, is prediction, the prediction of the incidence of the public force through the instrumentality of the courts." Id. at 457.

21. A paradigmatic expression of this classical ideal is Christopher Columbus Langdell's claim that "law is a science, and that all the available materials of that science are contained in printed books." Christopher C. Langdell, Harvard Celebration Speeches, *in* 3 L.Q. Rev. 123, 124 (1887).

For all of its conceptual impact on twentieth-century legal scholarship,[22] however, there is one methodological invitation quite literally offered by Holmesian prediction theory that legal scholars have generally not taken up—they have only rarely explored systematic methods of predicting the outcome of cases prospectively. Holmes was upfront about the limitations of his own formulation for actually predicting cases, disclaiming that "[t]heory is my subject, not practical details."[23] And his thin proposal for doing prediction was remarkably conventional: Study the "body of [case] reports, of treatises, and of statutes."[24] For the Realists who followed Holmes, it was likewise easier to theorize negatively against a prior generation's classical doctrine than it was to offer a new affirmative theory about how we might assess, predict, and discern regularity in judicial decisionmaking in a world where doctrine did not always constrain judges. That many Realists never offered much beyond judges' idiosyncratic "hunches" in terms of positive predictive theory was one of the movement's failings,[25] and one keenly recognized by Karl Llewellyn and others. Throughout his long career Llewellyn searched hard for general factors to aid in the prediction, or "reckonability," of court behavior—factors that were not linked to the particularities of case-specific doctrine or text.[26]

---

22. The basic proposition that judges exercise some degree of discretion in deciding cases has directly or indirectly motivated a great amount of legal and political science scholarship in the past century. Some of these questions sound in political theory, such as the longstanding debate in constitutional law over the alleged "countermajoritarian difficulty" posed by unelected judges who exercise meaningful authority. See generally Barry Friedman, The Birth of an Academic Obsession: The History of the Countermajoritarian Difficulty, Part Five, 112 Yale L.J. 153 (2002) (tracing evolution of academia's focus on the countermajoritarian difficulty and placing this focus within body of scholarship seeking to justify judicial review). Other questions are more pragmatic and empirical, and these indirectly motivate this study: *How much* discretion do judges have to choose among alternative outcomes, particularly at the Supreme Court, where such discretion is greatest? Do "legal" sources—such as precedent or legal text—constrain the Justices in a meaningful and recognizable way, and if so, how? Where text and precedent do not constrain, what other factors drive judicial decisionmaking? Are these nonlegal factors predictable and generalizable, or hopelessly idiosyncratic and personal? Do the Justices act differently in different kinds of cases with different doctrinal and institutional settings? And finally, what methods of assessing Supreme Court decisions best illuminate the foregoing queries?

23. Holmes, supra note 20, at 477. See also Frederick Schauer, Prediction and Particularity, 78 B.U. L. Rev. 773, 774 (1998) (describing significant theoretical commentary on Holmes's argument, but noting that "[m]uch less attention has been focused on the idea of prediction itself, or on the mechanisms by which a person . . . might predict what the law will do").

24. Holmes, supra note 20, at 457.

25. See generally Morton J. Horwitz, The Transformation of American Law 1870–1960, at 193–212 (1992) (discussing major tenets, strains, and legacy of Legal Realism).

26. See Karl N. Llewellyn, The Common Law Tradition: Deciding Appeals 17–18, 223, 335–36 (1960) [hereinafter Llewellyn, Common Law Tradition]; see also K.N. Llewellyn, On the Good, the True, the Beautiful, in Law, 9 U. Chi. L. Rev. 224, 243–46

With the waning of Realism in the law schools, much of the academic interest in prediction of cases shifted across campus to the fledgling field of quantitative political science as applied to courts. Spurred by the legal positivist impulse but also possessing specialized training in statistics, formal modeling, and other methodological tools—and unburdened by any strong commitment to lawyers' forms of thought and analysis—mid-century political scientists took research on judicial behavior in important new directions. Some of this work is characterized by an effort to reduce judicial decisionmaking to a few general explanatory variables, and then study a large number of court results (typically not opinions) to assess consistency with these factors.[27]

One political science model that arose early in this story and has remained prominent is the "attitudinal" model of Supreme Court behavior. In its purest form, this model posits that the Justices generally decide cases based upon their fixed policy preferences—that is, their personal ideological views—and are not meaningfully constrained from voting in accord with those views by doctrine, text, or institutional setting.[28] Moreover, in the standard attitudinal view the Justices are arrayed neatly along one or more linear dimensions based on a "liberal" to "conservative" spectrum of personal views (think of an abacus with nine beads), and most decisions track this ideological lineup. In quantitative studies run retrospectively, the attitudinal model has been very successful in accounting for—technically "postdicting"—the outcomes of Supreme Court cases.

For all of its postdictive success, however, there are a few problems—both technical and conceptual—with using the standard attitudinal model to predict cases. The technical problems are twofold. The first is that the attitudinal model is quite good at predicting the Justices' array along a particular linear dimension. But in its basic form it is not particularly good at situating specific cases ex ante along that linear array so as to predict *where* the key decision point will be—that is, how many Justices will vote one way and how many the other. As long as the Justices' votes align according to the predicted spatial array, the outcome is regarded as

---

(1942) (describing a "new style" of legal scholarship that is rooted in "conscious and overt concern" about policy, factuality, and scale); K.N. Llewellyn, On Reading and Using the Newer Jurisprudence, 40 Colum. L. Rev. 581, 587 (1940) ("The method is to take accepted doctrine, and check its words against its results, in the particular as in the large. . . . and to be content with no formulation which does not account for all of the results.").

27. We do not attempt here to summarize the various schools of thought about judicial behavior that exist in the modern political science academy, much less provide a detailed historical treatment. For good recent descriptions of these academic developments up through the present day, see Lee Epstein & Jack Knight, Toward a Strategic Revolution in Judicial Politics: A Look Back, a Look Ahead, 53 Pol. Res. Q. 625 *passim* (2000), and Michael Heise, The Past, Present, and Future of Empirical Legal Scholarship: Judicial Decision Making and the New Empiricism, 2002 U. Ill. L. Rev. 819, 833–43.

28. See Jeffrey A. Segal & Harold J. Spaeth, The Supreme Court and the Attitudinal Model 65 (1993).

consistent with the attitudinal model, irrespective of the decisional dividing line.[29] So for instance, on the current Court a unanimous decision either way is consistent with the attitudinal "prediction," but so too is a 5-4 decision where Justice O'Connor joins Rehnquist/Thomas/Scalia/Kennedy, and so too is a 5-4 decision where she joins the Stevens/Ginsburg/Breyer/Souter quartet. The only type of decision that flunks the spatial model is one where, say, Justices Scalia and Thomas vote with Stevens, Ginsburg and Souter to vacate a defendant's sentence and Justice Breyer is with Rehnquist, O'Connor and Kennedy in dissent.[30] Clearly, a model that would claim predictive accuracy in a case like *Grutter v. Bollinger*,[31] irrespective of whether Justice O'Connor voted to uphold or strike down the affirmative action plan at issue, leaves much to be desired.

There is another technical problem with using the standard attitudinal model to forecast cases prospectively: In the retrospective studies, postdiction for a specific case is achieved by matching up general variables with case-related factors contained in the Supreme Court opinions for that case. To predict cases using attitudinal models requires overlaying a host of case-specific variables onto the basic spatial array. But this increased accuracy often comes at a methodological price—the fact-specific variables are often more numerous than the number of cases predicted.[32] Jeffrey Segal solved some of these problems in a predictive study of search and seizure cases (again technically postdiction), but his successful effort there depended upon his review of the Court's major search and seizure decisions—that is, the cases to be explained—to identify the relevant variables.[33] Moreover, his study only identified variables useful for prediction in a narrow area of law (search and seizure cases). None of these efforts purport to be generally applicable across all of the Court's cases, or to apply without regard to case-specific facts.

Beyond these technical problems, there is a conceptual step that the leading proponents of this attitudinal model make that has generated skepticism about the value of their postdictive studies. The claim is that

---

29. On the current Court, this presumed spatial array has Stevens at one pole, followed in order by Ginsburg, Breyer, Souter, O'Connor, Kennedy, Rehnquist, Scalia, and Thomas. See Andrew D. Martin & Kevin M. Quinn, Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999, 10 Pol. Analysis 134 *passim* (2002). Although the Stevens end of the array is often labeled "liberal," and the Thomas end "conservative," spatial voting consistency is revealed by the Justices irrespective of the addition of such substantive labels for the opposite poles.

30. See, e.g., Apprendi v. New Jersey, 530 U.S. 466 (2000).

31. 123 S. Ct. 2325 (2003).

32. See, e.g., Fred Kort, Predicting Supreme Court Cases Mathematically: A Quantitative Analysis of the "Right to Counsel" Cases, 51 Am. Pol. Sci. Rev. 1, 4–6 (1957).

33. See Jeffrey A. Segal, Predicting Supreme Court Cases Probabilistically: The Search and Seizure Cases, 1962–1981, 78 Am. J. Pol. Sci. 891, 892–93 (1984). The same is true for some of the rare legal academic forays into prospective forecasting. Fred Rodell's nonquantitative prediction exercise in 1962 was accurate, but predicted only one case, *Baker v. Carr*. See Fred Rodell, For Every Justice, Judicial Deference Is a Sometime Thing, 50 Geo. L.J. 700, 707–08 (1962).

the liberal-to-conservative preference array does not merely *correlate* with judicial voting patterns, but that it is the primary *cause* of those votes. As the two leading proponents of the approach put it a decade ago: "Rehnquist votes the way he does because he is extremely conservative; Marshall voted the way he did because he is extremely liberal."[34] This general pattern may hold, but the Justices' votes and ideologies are not necessarily linked precisely in this causal way. Other factors may make the decisional process more complex and nuanced than the attitudinalists' account, and many legal and political science skeptics have made these points. In its major form—"judicial ideology is all that matters"—attitudinalism is pilloried for claiming too much; in its minor form—"judicial ideology matters sometimes"—it is dismissed as telling us what we already knew, and with lots of unpleasant counting of cases to boot.[35]

It is possible to broaden this kind of critique beyond attitudinalism to encompass almost any general theory of Supreme Court decisionmaking. Do judicial attitudes, and institutional setting, and doctrine and text, and broad principle and history matter to the Court's outcomes? Almost certainly yes, yes, yes, and yes. Does any one or two of these factors explain everything? Probably not. We think it probable that all of these factors (and more) contribute in one way or another to the choices that Justices make. If decisionmaking is multifactorial in this sense, then it should not be surprising that analytically deft legal academics and political scientists can find evidence of their preferred factors in the Court's past behavior and write persuasive scholarship advancing their views. The problem is not that this diverse scholarship is defective, but rather that it is so successful at advancing—within the analytical frameworks acceptable in each discipline—a myriad of different factors that probably correlate with judicial choices *to a greater or lesser extent in individual cases*. This last qualifier, however, is critical in actual prediction, for it is precisely this greater or lesser degree to which various factors matter in real cases that lead to real outcomes.

It is here that prospective prediction experiments can be helpful, not necessarily by directly proving or disproving underlying causation, but by measuring and assessing how various factors correlate with actual decisionmaking in different kinds of cases. One clear benefit of predictive efforts is that their success is verifiable or refutable with the passage of time in a way that retrospective analytical work is not. Prediction exercises thus have the potential to revise or unsettle preexisting academic attitudes in ways that retrospective analyses of past data may not. Although mere prediction does not itself prove causation, the exercise of

34. Segal & Spaeth, supra note 28, at 65.

35. Lon Fuller disparaged quantitative research on judicial behavior on these grounds in 1966, writing that it adds "[n]ot much by way of practical utility" and that it was an inefficient "scientific enterprise that seems to return so little from so much." Lon L. Fuller, An Afterword: Science and the Judicial Process, 79 Harv. L. Rev. 1604, 1622 (1966).

constructing and testing predictive models can advance both explanation and understanding.[36]

II. Project Design and Methodology

The purpose of this study is to compare two different ways of assessing and forecasting Supreme Court decisionmaking. Its basic structure necessarily shapes—and in some ways limits—our findings. In this Part, we first explain certain methodological choices in the overall design of the study, then describe in greater detail the statistical model, and finally, explore the nature of the legal experts' decisionmaking.

A. *Overall Study Design*

1. *The Rehnquist Supreme Court.* — We chose to run our comparative study by focusing on a single court—the United States Supreme Court—and its output in a single Term.[37] The Supreme Court is an obvious object of study, both because of its importance as an institution and because of the wealth of analytical and empirical scholarship and objective data that have been collected regarding its work. Moreover, *this* Supreme Court offers a unique opportunity for research because the same nine Justices had been sitting together for nearly a decade prior to the 2002 Term. Because of the longevity of this natural court,[38] both the statistical

---

36. Broad predictive exercises on the Court such as this are rare, although not unprecedented. Harold Spaeth in the 1970s predicted several dozen selected Supreme Court cases per year, often with high success rates. See William K. Stevens, The Professor's Computer Foretells Court's Rulings, N.Y. Times, July 28, 1974, at 41. There appears to be a general increase in interest in prediction both specifically of the Supreme Court and in many other settings. Attorney Sam Heldman predicted the outcomes of every case on the Supreme Court's 2002 Term docket on his legal weblog. See Sam Heldman, Ignatz: Law and Politics (June 27, 2003), at http://sheldman.blogspot.com/2003_06_01_sheldman_archive.html (on file with the *Columbia Law Review*). Another website promoted and ran a contest entitled "Supreme Court Fantasy League" for predictions of selected cases in the 2002 Term and is currently running another for the 2003 Term. See http://www.lawpsided.com/lawpsidedcontests.htm (last visited Feb. 25, 2004). Recent months have seen a more general academic and popular interest in prediction methodology in fields as diverse as baseball and world terrorism. See, e.g., Richard H. Thaler & Cass R. Sunstein, Who's on First, The New Republic, Sept. 1, 2003, at 27 (reviewing Michael Lewis's *Moneyball*, a book about innovative statistical techniques for forecasting baseball performance developed by statistician Bill James and applied by the Oakland A's and other teams); Michael Abramowicz, Information Markets, Administrative Decisionmaking, and Predictive Cost-Benefit Analysis, 71 U. Chi. L. Rev. (forthcoming Summer 2004) (describing the ill-fated proposal by the Defense Advanced Research Projects Agency to develop a "terrorism futures market").

37. We included all argued cases in the Court's regular October–June 2002 Term. We did not include in our analysis the campaign finance case argued on September 8, 2003 (McConnell v. FEC, 124 S. Ct. 619 (2003)), even though that case was technically argued during the October 2002 Term. See Sup. Ct. R. 3.

38. We adopt the commonly accepted definition of "natural court" as referring to a period of time where the same nine Justices sit together on the Supreme Court without any composition change. See, e.g., Joan Biskupic & Elder Witt, The Supreme Court at Work

model and the legal experts have the benefit of hundreds of cases decided by these same nine individuals on which to base predictions about their future behavior.

2. *Method, Not Theory*. — Our study compares different methods of prediction. It does not directly contrast two mutually exclusive theories about what motivates the Court. Although scholars who study the Court have long debated the motivations underlying the Justices' decisions, we do not join the stylized debate between "legalism" and "attitudinalism" in any precise sense. Neither of our methods of prediction is designed to test a pure theory of what motivates the Justices—indeed, the individual legal experts considered both legal and nonlegal factors in reaching their predictions,[39] and the variables utilized by the model do not capture solely ideological motivations. Still, in ways we describe more fully below, underlying theoretical differences separate the two prediction methods. The model used inputs derived in significant part from decades of political science research on judicial decisionmaking that often began with attitudinal assumptions. Conversely, although our legal experts were not strictly limited to considering only "the law," they were chosen because of their expertise in thinking about and writing about legal doctrine.

Despite this theoretical divergence, the most essential contrast between the two methods we employ lies in the differing nature of the inputs used to generate predictions. The statistical model took into account the outcome of all 628 cases decided by this natural court prior to the October 2002 Term. In doing so, it gave each of those cases equal weight in constructing the classification trees used to generate its predictions. The machine also relied on only a handful of characteristics about those cases, each of them gross features easily observable without specialized training. Although those characteristics might serve as proxies for important aspects of the legal process, they are inherently blind to specific legal doctrines and texts.

By contrast, the legal experts were unlikely to consider all of the Court's decisions over the prior eight terms in reaching their predictions. The nature of legal study—focused as it is on leading cases—predisposes legal experts to focus on a handful of salient cases, rather than attempt to weight all cases equally. Even if they wanted to, basic cognitive limitations would prevent the human experts from systematically and equivalently taking account of every case previously decided by this natural court. However, unlike the machine, the legal experts could recognize and take account of particularized knowledge such as the facts of the case, specific

---

315 n.a (2d ed. 1997). Scholars have taken to referring to the current Court's longstanding membership stability since October 1994 as the "second Rehnquist Court." See Thomas W. Merrill, Childress Lecture: The Making of the Second Rehnquist Court: A Preliminary Analysis, 47 St. Louis L.J. 569, 570 (2003).

39. Few legal experts today are likely to be pure "legalists," who would base prediction and analysis exclusively on neutral doctrine and text without any inquiry into the particular composition of the Court.

legal doctrine or texts, or statements by individual Justices in similar cases. Thus, the experts, as compared with the machine, relied on fewer, but more detailed, observations of past Court behavior.

3. *Outcomes, Not Opinions.* — In comparing the two methods, we focus on the *outcomes* of Supreme Court cases, not their internal content. We designed the machine, and asked the experts, to make only a binary choice between affirm or reverse outcomes in the Term's cases. We acknowledge that such a binary choice offers an incomplete picture of the Court's work, but defend this focus on both substantive and methodological grounds. First, the basic outcomes produced by the Court impact American society profoundly in ways that transcend the specific rationales offered by the Justices. Legal scholars continue to debate and critique the judicial rationales offered in crucial cases such as *Brown v. Board of Education*,[40] *Roe v. Wade*,[41] and *Bush v. Gore*,[42] but for most of the nation's citizens it was the basic outcome of those decisions that carried the most weight, and continues to do so. More recently, in the case of *Lawrence v. Texas*,[43] the distinction in legal reasoning between Justice Kennedy's majority opinion and Justice O'Connor's concurrence is interesting and important, but for most Americans this distinction pales in comparison to the essential fact that six Justices declared unconstitutional the Texas prohibition on consensual homosexual sodomy.

Second, and more pragmatically, outcomes provide a common ground on which to compare the predictive performance of the legal experts and the machine. Although lawyers can and do make predictions about both outcomes and reasoning, the model is incapable of generating predictions about the content of the Court's opinions. By design, the statistical model is blind to legal doctrine in its inputs—and thus, it is correspondingly mute as to doctrine in its predictive outputs. In this sense, the human experts have a broader analytical skill set, and one that is vastly underutilized in this study. But in order to have a uniform point of comparison between the two methods, we needed to restrict our focus to outcomes, the only type of prediction the model could produce.

None of this is intended to say that internal opinion content is unimportant, for the Justices' rationales undoubtedly affect lower courts and future legal developments in critical ways. The reasons the Justices give for their opinions matter, whether or not one regards the reasons given as a complete explanation of behavior. The Court's opinions provide the rules that lower courts apply, constitute the object of scholarly commentary and critique, and shape public discourse on important issues. Because our study does not account for this content, there is much it does not, and cannot, say about the judicial process. We readily acknowledge the limitations of a study, like ours, that would have treated the most

40. 347 U.S. 483 (1954).
41. 410 U.S. 113 (1973).
42. 531 U.S. 98 (2000).
43. 123 S. Ct. 2472 (2003).

famous case in American history as simply "Marbury loses," without any concern for what John Marshall actually said in reaching that result.[44] But such a limitation is both substantively defensible and methodologically necessary in this sort of comparative study.

B. *The Statistical Model*

Our principal goal in constructing the statistical model was to create a computer program capable of predicting the outcome of Supreme Court cases prospectively, using only information available prior to oral argument. For reasons explained below, we used classification trees for the statistical forecasting model. The model's predictions depended on only six variables: (1) circuit of origin; (2) issue area of the case; (3) type of petitioner (e.g., the United States, an employer, etc.); (4) type of respondent; (5) ideological direction (liberal or conservative) of the lower court ruling; and (6) whether the petitioner argued that a law or practice is unconstitutional.[45] This information, when fed into the classification trees, generated a predicted vote for each Justice and a predicted outcome for each case pending before the Court in the 2002 Term.[46]

In creating the statistical model, we began with an assumption of temporal stability in the Justices' behavior. In other words, we assumed that observable patterns in the Justices' past behavior would hold true for their future behavior. In order to capture these patterns, we utilized data from all 628 cases decided by this natural court prior to the October 2002 Term, which we refer to as our "training data." We selected a number of variables plausibly correlated with outcomes for potential inclusion in the model, of which six were incorporated in the final model.[47]

Because our goal was predictive accuracy, not hypothesis testing, no formal theory of Supreme Court decisionmaking drove our choice of vari-

---

44. See Marbury v. Madison, 5 U.S. (1 Cranch) 137 (1803).

45. "Circuit of origin" includes cases on appeal from a state or a three-judge federal district court panel located within a particular circuit. "Issue area" corresponds to the VALUE variable in Harold Spaeth's Supreme Court database. See Harold J. Spaeth, The Original United States Supreme Court Judicial Database, 1953–2002 Terms, Documentation 51 (last updated Nov. 25, 2003), available at http://polisci.msu.edu/pljp/sctcode.PDF (on file with the *Columbia Law Review*) [hereinafter Spaeth, Documentation] (explaining definition of VALUE variable). "Type of petitioner" and "type of respondent" also used Spaeth's coding protocol, but several categories were collapsed. For cases pending in the 2002 Term, all six variables were coded from the petitioners' merits briefs before the Supreme Court using Spaeth's coding protocol.

46. The model generates predicted probabilities for each possible outcome. If the forecasted probability of a reversal is greater than 50%, it is treated as a simple "reverse" prediction, and likewise for predicted affirmances.

47. The variables considered for inclusion were: liberal or conservative direction of the lower court decision, issue area, circuit of origin, identity of the petitioner, identity of the respondent, argument that a practice is unconstitutional, manner in which the Court took jurisdiction, petitioner claim of lower court disagreement, whether the case came from a state supreme court, and whether the petitioner argued that the Court should overturn precedent. Only the first six of these variables were used in the final model.

ables. Rather than searching for ideal measures of some explanatory variable, we relied largely on pragmatic considerations. In order to be useful for forecasting, the information had to be easily observable and readily available *prior* to oral argument. We excluded some potentially useful variables simply because the information was too difficult to collect. Nevertheless, the selection of potential variables drew on existing literature about the Court, and, in particular, attitudinalist insights. For example, attitudinal models of Supreme Court voting suggest that whether the lower court decision was liberal or conservative will often correlate (positively or negatively) with the votes of the Justices. Similarly, the identity of the parties might affect the political valence of a case, and philosophical or ideological differences between the circuits might lead to differing patterns of responses from the Justices. But although they influenced our choice of potential variables, basic attitudinal assumptions are insufficient to generate specific forecasts prospectively for the reasons discussed in Part I. The basic attitudinal model fails to specify ex ante where a particular case will fall along its predicted linear array. Without some kind of leverage on the case facts, the model cannot generate predictions prospectively.

Because of this limitation of spatial voting models, we turned to classification tree analysis as a way to generate predictions from case-specific information.[48] Classification trees have been used in other contexts for forecasting, and provide a flexible method for pattern finding in situations involving many variables. They enabled us to capture patterns in the Justices' observable past behavior without assuming a linear relationship between covariates and outcomes. Using all of the potential variables and information about actual outcomes in the training data, we estimated the classification trees that best fit the past cases.[49] Interestingly,

---

48. There are other technologies that could be used to forecast Supreme Court behavior. One of particular note is the use of neural network models. Our choice to use classification trees is motivated by the transparency of the model; i.e., trees are produced that can be graphically represented and easily studied. See Appendix A. Other approaches tend to be more of a "black box," and, as such, are very difficult to understand.

49. In brief, we started with a set of twenty-four potential models. These models differed from one another based on our choice of certain parameters—for example, whether unanimous cases were forecast separately or not, and whether the individual Justices' votes were linked or independent. We then split the pre-2002 data into two mutually exclusive parts, which we refer to as the in-sample data and the out-of-sample data.

For each potential model specification, we fit the model to the in-sample data, used it to predict the out-of-sample decisions, and calculated the percentage that were correctly predicted. We chose as our forecasting model the model specification that did the best job of classifying the out-of-sample decisions. Thus, the model selected was the one that maximized correct case outcome predictions, not correct vote predictions. Finally, we fit this model to the full pre-2002 data and then used it to forecast decisions in the October 2002 Term. Although it would be possible to assign probabilities to different outcomes (e.g., a 60% chance of affirmance), we treated the model's forecasts as all having a probability of one. In other words, the model's forecasts, like the experts', were captured

the final classification trees did not utilize all of the variables initially selected. Some simply "dropped out" of the trees, having no predictive power.[50] The six variables included in the model were retained simply because they best fit the training data.

For each case pending before the Supreme Court during the 2002 Term, we coded the six variables used by the model and used them to generate the machine's forecasts. The final model consisted of eleven distinct classification trees. The first two predict whether a case is likely to be a unanimous "liberal" decision or a unanimous "conservative" decision. These two trees were applied first for every case prediction, and the process ended there if a unanimous result in one direction was predicted. However, if neither of the first two trees predicted a unanimous decision (or if *both* did, in opposite directions), then nine separate classification trees—one to forecast the vote of each Justice—were utilized. As an example, Figure 1 presents the estimated classification tree that was used to forecast the votes of Justice O'Connor. Consider *Grutter v. Bollinger*,[51] the case challenging the constitutionality of Michigan Law School's affirmative action policy. Proceeding to the first decision point in O'Connor's tree, the model (erroneously) forecasts a reversal because the lower court decision was liberal.[52] Had the lower court decision been conservative, the case would have dropped to the next branch, which asked which circuit the case was from, and continued in like manner down the tree until a final prediction emerged for O'Connor's vote.

Appendix A contains diagrams of all eleven classification trees in the statistical model. The structures of these trees are interesting independent of the outcome of the forecasting exercise. A quick visual comparison reveals that the trees vary significantly from Justice to Justice. Not only do they differ in terms of their overall shape and the number of branches they contain, but variables figuring prominently in the decision tree of one Justice may be relatively unimportant or altogether absent in another.

Some of the Justice-specific classification trees take into account the predicted votes of other Justices. For example, the statistical model's forecast of Justice Breyer's vote depends on the predicted votes of Justices O'Connor and Ginsburg in the same case,[53] requiring that those two

---

as simple "affirm" or "reverse" predictions without attempting to assess the probability of each possible outcome.

The twenty-four models we tested obviously do not exhaust the universe of possible models. Choosing different model parameters might produce more accurate forecasts. Nevertheless, we decided to test a reasonable number of models, all of them plausible on substantive grounds. Our final statistical model is a product of those choices.

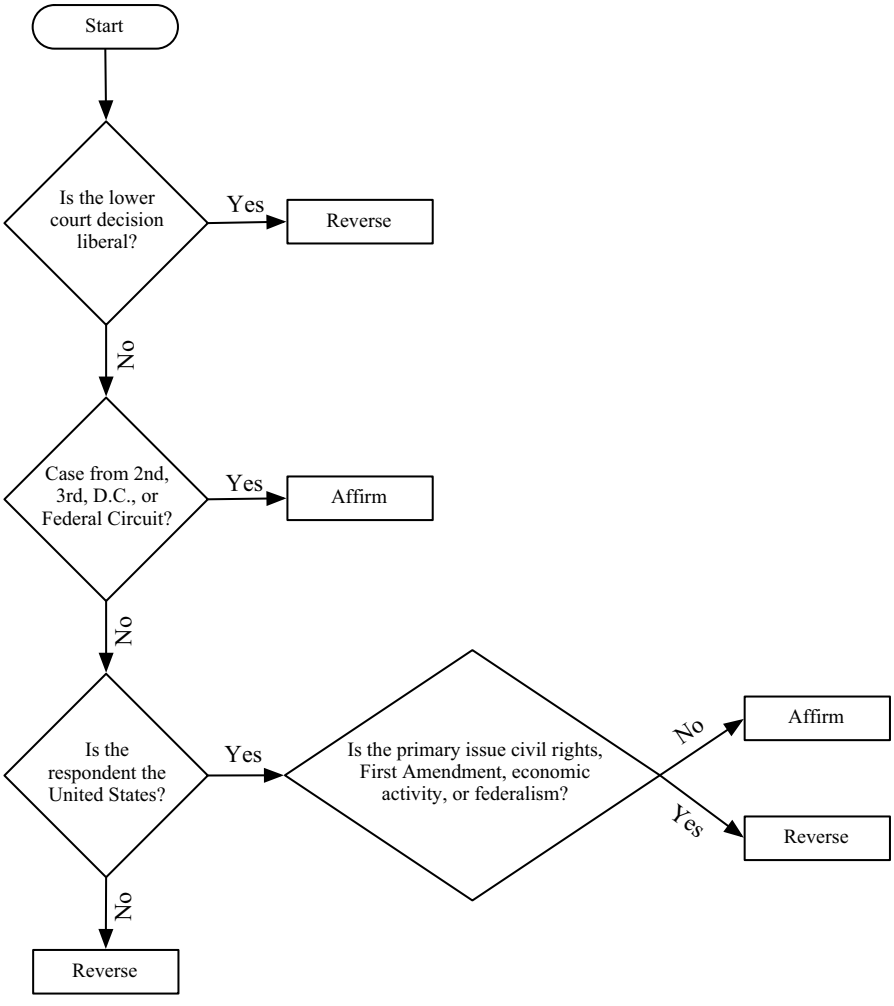
50. See supra note 47 (listing the ten variables considered, including the four that were ultimately discarded).

51. 123 S. Ct. 2325 (2003).

52. Grutter v. Bollinger, 288 F.3d 732 (6th Cir. 2002).

53. The relationships between the Justices that are visible in the classification trees generate reasonably good predictions of the Justices' votes because they capture

FIGURE 1: ESTIMATED CLASSIFICATION TREE FOR JUSTICE O'CONNOR FOR FORECASTED NON-UNANIMOUS CASES.



votes be predicted first. Accordingly, the model generates predictions of the Justices' votes in a precise sequential order: unanimous liberal outcome, unanimous conservative outcome, Scalia, Thomas, Rehnquist, Stevens, O'Connor, Ginsburg, Breyer, Souter, and then Kennedy. Thus, Justice Scalia's predicted vote on a case is generated before Justice Thomas's, whose predicted vote might vary based on the Scalia predic-

correlations in their behavior. They should not be interpreted as claiming that one Justice's vote causes or motivates the behavior of another.

tion.[54] The model then uses the predicted vote of Justice Thomas as a relevant variable in generating Justice Kennedy's predicted outcome.[55]

Prior to oral argument, we posted the statistical model's predictions for each case on a project website.[56] At the end of the Term, however, we became aware of an error in the software code that was used to input the case characteristics into the model. As a result, several of the forecasts originally posted on the website did not actually reflect the operation of the model as written. The programming error had the effect of misclassifying some cases as they were run through the model's decision trees, so that, for example, a case from the Ninth Circuit was erroneously entered into the model as if it had originated in the Fifth Circuit. We corrected that programming bug and regenerated all of the machine's forecasts.[57] As a result of correcting the programming error, the machine's overall predictive accuracy improved from 68% to 75%, but otherwise the basic results were unchanged.[58] This Essay reports and analyzes the forecasts generated by the model operating with the corrected software. We believe, however, that the potential for a programming error to affect outputs of machine-based prediction is itself worthy of note. In this sense, a stark contrast between "humans" and "a machine" is misleading. The machine is itself a product of a series of choices made by humans, including which variables to consider for inclusion, how to code them in particular cases, and how to use them to generate outcomes. The efficacy of the machine ultimately depends on those human choices and remains vulnerable to them and the risk of human error.

C. *The Legal Experts*

The study's other method of prediction seeks to capture the case-specific judgments of a large number of legal experts. Experts are distinguished from nonexperts by extensive training and experience in the relevant domain. In addition to greater specialized knowledge, experts have the ability to perceive meaningful patterns and to structure their knowledge on deeper, principle-based categories. Often, their judgments are

---

54. See Figures 8 and 9 in Appendix A.

55. See Figure 16 in Appendix A.

56. See The Washington University Supreme Court Forecasting Project, at http://wusct.wustl.edu (last updated Apr. 5, 2004) [hereinafter Project Website] (providing data on the 2002 forecasts, as well as forecasts for the current 2003 Term).

57. The statistical model itself—that is, the classification trees—had not changed since the start of the 2002 Term, nor had the manner in which the characteristics of the pending cases were coded. In most cases, the machine's forecasts did not change, although in seven cases the machine's prediction switched from "affirm" to "reverse" when the error was corrected, and in six cases the opposite occurred.

58. The website for this project includes a list of all the old, incorrect forecasts, as well as more technical information about the programming error and computer code that can be used to replicate all of the forecasts. See Project Website, supra note 56 (click on hyperlinks available at http://wusct.wustl.edu/2002/errors/index.html).

based on analyses that are qualitative in nature. In each of these ways, the judgments of our legal experts differed from that of the machine.

Because no metric exists to measure expertise precisely, we recruited participants much the way anyone might look for expert assistance: We researched their writings, checked their training and experience, and relied on our own personal knowledge and referrals from knowledgeable colleagues in their fields. The eighty-three individuals who participated comfortably qualify as "experts," having written and taught about, practiced before, and/or clerked at the Court, and having developed significant expertise in one or more substantive fields of law. Collectively, they form an accomplished group of seventy-one academics and twelve appellate attorneys, comprised of thirty-eight former Supreme Court law clerks, thirty-three chaired professors, and five current or former law school deans. The names of the participating experts are listed alphabetically in Appendix B. We note with much gratitude that the experts' participation was an entirely volunteer effort, and their substantial intellectual generosity made this project possible.

We asked experts to predict a case or cases within their areas of substantive expertise.[59] More than one expert predicted most cases, but experts assigned to the same case did not communicate about their predictions and were unaware of one another's identity. We requested their forecasts prior to oral argument, and assured them that we would not reveal their individual predictions or the cases to which they were assigned. Experts were free to consider any sources of information or factors they thought relevant to making their prediction.[60] In addition to an "affirm" or "reverse" prediction for the Court as a whole and for each Justice,[61] some experts also offered brief written comments about the case or their prediction.

---

59. To best match particular cases with particular expertise, and to avoid overburdening our volunteer experts, we limited each participating expert to predicting between one and three cases. One expert predicted four cases. We matched experts with cases using an "issue preference form" that the experts completed.

60. We provided a copy of the lower court opinion and citations to the parties' Supreme Court briefs, but did not limit the experts to these materials.

61. For those inclined to parse different legal questions differently (as most legal academics and lawyers are), the requirement of a single "affirm" or "reverse" prediction seems unrealistically simplistic. Although this artificial bluntness understandably frustrated some experts, we do not think it necessarily affected the comparative results. Forcing a single binary choice essentially required the experts to decide which issue they thought would be crucial to the Court's decision, and to base their prediction in the case on the outcome of that issue. Some expert predictions might have been incorrect because they misapprehended which issue would be crucial, even though they would have made a correct prediction on another issue. However, the model would be equally if not more vulnerable to this risk, as it bases its predictions on general trends without any regard to the specifics of the case. And in some cases the experts could, and did, recognize specific grounds for decision that were so particularized (and often technical) as to be absolutely beyond the machine's recognition. See discussion infra Parts III–IV.

By asking a group of legal experts to make predictions, we did not expect to capture a single coherent theory of Supreme Court decisionmaking. Our experts are diverse in their experiences, areas of expertise and philosophies. We fully expected that they would differ in the factors they thought important to consider, and in how they applied them in particular cases. Instead, what we sought was the best judgment of individuals with legal expertise—that is, those with the training and knowledge to take account of specific legal factors, such as doctrine or text, to the extent they thought appropriate, along with whatever other factors they deemed relevant. Although another group of experts might have approached this task differently—and perhaps produced different results—this group certainly had the capacity and experience to assess meaningfully a host of legal and nonlegal variables in making their predictions.

Although it is impossible to trace precisely how the experts reached their predictions, we obtained some information about the factors that played a role in their decisionmaking process. Upon receipt of an expert's prediction in a particular case, we sent that expert a written survey asking him or her to rate a list of factors that were important to his or her prediction.[62] The survey responses, together with their written comments, offer a glimpse of how one cross-section of legal experts perceives Supreme Court decisionmaking.[63]

D. *The Court's Decisions*

Throughout the Term, we posted all of the machine and expert forecasts prior to oral argument on the project website.[64] After each decision, we coded the actual outcome in each case and the vote of each Justice as "affirm" or "reverse." In doing so, we focused on the bottom line outcomes: Cases that were vacated and remanded, or reversed even in part, were coded as "reverse."[65] Concurring votes—even ones that dif-

---

62. Experts predicting more than one case received a survey for each case. In all, approximately 90% of our experts returned at least one survey, and we received responses for 65% of the expert predictions made during the Term. In order not to influence the experts' predictions by exposing them to the survey's list of factors potentially influencing the Court's decision, we sent the surveys to each expert after receiving his or her prediction in a particular case. The downside to this choice was that it required experts to recall their decisionmaking process after a week or two had passed.

63. Of course, these data are not direct evidence of their thought processes. Problems of recall or unconscious biases might affect the accuracy of our experts' self-reports. Nevertheless, some interesting patterns emerge from what the experts say were the factors that influenced their predictions.

64. See Project Website, supra note 56. As discussed supra Part II.B, some of the machine's forecasts posted during the Term were incorrect due to a software error. The analysis reported here uses the corrected forecasts.

65. Obviously, not all reversals are equal from the perspective of future litigants or even the parties themselves. When the Supreme Court reverses and remands on narrow grounds, the petitioner may win temporarily but end up losing the case after the new legal standard is applied. For example, in Sell v. United States, 123 S. Ct. 2174 (2003), the Court

fered dramatically in terms of rationale—were treated the same. For instance, Justice O'Connor's vote in *Lawrence v. Texas* was coded "reverse," just like the votes of the five Justices joining the majority opinion, even though she advocated reversal on quite different grounds.[66]

Using these criteria, the coding decision was straightforward in most cases. However, we excluded several cases in which no opinion was issued, or for which the outcome could not fairly be characterized in simple "affirm" or "reverse" terms.[67] In all, we used sixty-eight cases to analyze the case outcome forecasts and sixty-seven to analyze individual vote predictions.[68] Appendix C lists the machine and expert predictions and the actual outcomes for some of the major cases last Term. Predictions and outcomes for all of the cases included in our analysis are available on the project website.[69]

---

reversed the Eighth Circuit order permitting Sell's involuntary medication to render him competent to stand trial, but it did not prohibit such practices outright; the Court's decision left room for the government to try again on remand in accordance with the factors the Court announced. See id. at 2187. Despite the fact that Sell did not get the blanket prohibition he sought, we focus on the result at the Supreme Court level and code the outcome "reverse."

66. See 123 S. Ct. 2472, 2484 (2003) (O'Connor, J., concurring) (agreeing with the Court as to result but grounding her rationale in the Equal Protection Clause and not the Due Process Clause).

67. Of the seventy-six cases in which the Court heard oral argument, we excluded eight from our analysis. We excluded three cases because they were dismissed without opinion, see Nike, Inc. v. Kasky, 123 S. Ct. 2554 (2003); Abdur'Rahman v. Bell, 537 U.S. 88 (2002); Ford Motor Co. v. McCauley, 537 U.S. 1 (2002), and two because they were affirmed by an evenly divided Court, with no information about individual votes, see Dow Chem. Co. v. Stephenson, 124 S. Ct. 429 (2003); Borden Ranch P'ship v. United States Army Corps of Eng'rs, 537 U.S. 99 (2002).

We excluded three additional cases due to intractable coding ambiguities. Virginia v. Black, 123 S. Ct. 1536 (2003), involved several different defendants and substantive issues. Because different majorities of the Justices affirmed and reversed on the different issues, the case as a whole is impossible to categorize as either an "affirm" or "reverse." We also excluded Green Tree Financial Corp. v. Bazzle, 123 S. Ct. 2402 (2003), and National Park Hospitality Ass'n v. Department of Interior, 123 S. Ct. 2026 (2003), because in each case, the Court's decision turned on a preliminary issue. In *Green Tree*, the Court vacated and remanded, stating that whether the arbitration agreement permitted class arbitration must first be resolved by the arbitrator. See 123 S. Ct. at 2405. In *National Park Hospitality*, the Court also vacated and remanded, holding that the controversy was not yet ripe for judicial resolution. See 123 S. Ct. at 2028. Although technically each decision would be a "reversal" under our definition, the import of these decisions favored the respondents' positions, such that neither "affirm" nor "reverse" accurately captures the true outcome.

68. We excluded Chavez v. Martinez, 123 S. Ct. 1994 (2003), from our vote analysis only. In that case, coding the votes of individual Justices is impossible due to strategic concurrences (to form a Court judgment to vacate and remand) by Justices who stated that their substantive position was to affirm. No matter how we treat these ambiguous votes, the overall "reverse" outcome of the case is not in question, so we *do* include *Chavez* in our outcome analysis. We do not include it in our vote analysis, leaving sixty-seven cases where we summarize results for individual votes.

69. See Project Website, supra note 56.

III. Results

Comparing the accuracy of the two methods, the statistical model clearly did better than the legal experts in predicting case outcomes. In this section, we explain and analyze this basic outcome, breaking down the results by Justice, by issue area, and by type of legal expert. Part IV explores the implications of these results in greater detail.

A. *The Basic Results: This Round to the Machine*

The statistical model substantially outperformed the legal experts in forecasting case outcomes in the 2002 Term. As seen in Table 1, the machine correctly forecast 75% of ultimate case outcomes, while the experts' predictions were accurate only 59.1% of the time.[70] This difference between the machine and the experts in forecasting outcomes across all kinds of cases is statistically significant, even given the relatively small number of cases in the sample. A different result might well obtain in a different Term with the same or a different group of experts, but for this set of cases, the statistical model clearly performed better than the experts.

Table 1: Machine and expert forecasts of case outcomes for decided cases (N=68). Row percentages are in parentheses. The estimated (conditional maximum likelihood) odds ratio is 2.073 (P=0.025, Fisher's Exact Test).

	Case Outcome Forecast		
	Correct	Incorrect	Total
Machine	51 (75.0%)	17 (25.0%)	68 (100.0%)
Experts	101 (59.1%)	70 (40.9%)	171 (100.0%)

Table 1 treats each expert independently, summarizing the results by aggregating all available expert predictions. However, we had three experts predict most of the cases, with a view to isolating outlier expert predictions and capturing a majority, or consensus, expert prediction on most cases. Thus, an alternative measure of the experts' success takes the predictions of the majority of experts on a particular case as the experts' consensus prediction. Use of this measure improved the experts' success

70. Several of the expert forecasts were ambiguous due to narrative comments written on the ballot indicating different predictions on different legal issues in the case, or specifying that the prediction only applied to a single issue in a multi-issue case. We coded these ballots according to the first written prediction on the ballot, and included them in the reported results. We also performed the analysis without including these predictions. The substantive results are not affected. For reasons we explained in note 59, we believe that forcing the experts to make a single binary choice as to outcomes was unlikely to bias the results vis-à-vis the model.

rate somewhat. In the cases with a unanimous or majority consensus result, the experts' accuracy rate was 65.6%.[71]

Using the consensus predictions narrows the gap between the two methods,[72] but the basic result—the statistical model outperforms the experts—still obtains. The fact that the consensus predictions do not close the gap suggests that the experts' lower accuracy rate compared with the machine is not attributable to a handful of idiosyncratic expert predictions. Rather, the different success rates likely reflect systematic differences in the two methods of prediction—differences which we explore in greater detail in Part IV.

We also compared the success of the two methods in the Term's thirty-one unanimous cases. Many of the Court's closely divided cases involve ambiguous text and doctrine or divisive policy issues—it is not surprising that legal experts would have a hard time predicting those. But if some cases are decided unanimously because the relevant law is more determinate, then we might expect that experts trained to analyze legal arguments would outperform a model that is indifferent to specific doctrine and text. That did not happen with the Term's unanimous cases: Although the experts' success rate increased—to 65.3%—it remained behind the machine's 74.2%.[73] The fact that the machine's accuracy rate was marginally less, and the experts' only slightly greater in these cases suggests that the unanimous decisions that in hindsight look like "easy cases" are not obviously predictable prospectively.

B. *Predicting the Justices Who Matter Most*

Whether comparing aggregate expert predictions, consensus predictions, or only forecasts in the unanimous cases, the statistical model consistently outperformed the legal experts in predicting case outcomes. Perhaps surprising, then, is the fact that the model did slightly *worse* than the experts at forecasting the specific votes of the Justices in the Term's cases. Table 2 illustrates that the experts correctly predicted 67.9% of the Justices' individual votes during the Term, while the model lagged a bit

---

71. Cases with only two experts with opposite predictions were inconclusive. We excluded these cases altogether in calculating the 65.6% accuracy rate. Alternatively, we could have treated these inconclusives as incorrect (resulting in a 58.8% success rate for the experts) or assumed that if a third prediction had been obtained, the distribution of correct predictions would mirror the overall distribution of correct expert predictions (resulting in a 64.7% success rate).

72. The Condorcet Jury Theorem suggests that, on average, expert opinion aggregated in this fashion will outperform individual predictions. Consistent with the Theorem, the data suggest that experts do better when their votes are aggregated. Presumably, aggregating the predictions of greater numbers of legal experts would produce even better results. But arguably, the expert predictions should be treated independently, as the machine, at least in this project, was limited to a single predictive iteration.

73. This variance between the model and the experts is not statistically significant given the small number of unanimous results—thirty-one.

behind at 66.7%. Although the model and the experts did about equally well in predicting individual votes overall, on this Supreme Court not all votes are of equal importance in determining outcomes. Because the model did particularly well in predicting the centrist Justices who matter the most, it did significantly better at forecasting case outcomes.

TABLE 2: MACHINE AND EXPERT FORECASTS OF JUSTICE VOTES FOR DECIDED CASES (N=67). ROW PERCENTAGES ARE IN PARENTHESES. SOME JUSTICES DID NOT VOTE ON SOME CASES, AND ARE THUS NOT INCLUDED. THE ESTIMATED (CONDITIONAL MAXIMUM LIKELIHOOD) ODDS RATIO IS 0.943 (P=0.571, FISHER'S EXACT TEST).

	Justice Vote Forecast		
	Correct	Incorrect	Total
Machine	400 (66.7%)	200 (33.3%)	600 (100.0%)
Experts	1015 (67.9%)	479 (32.1%)	1494 (100.0%)

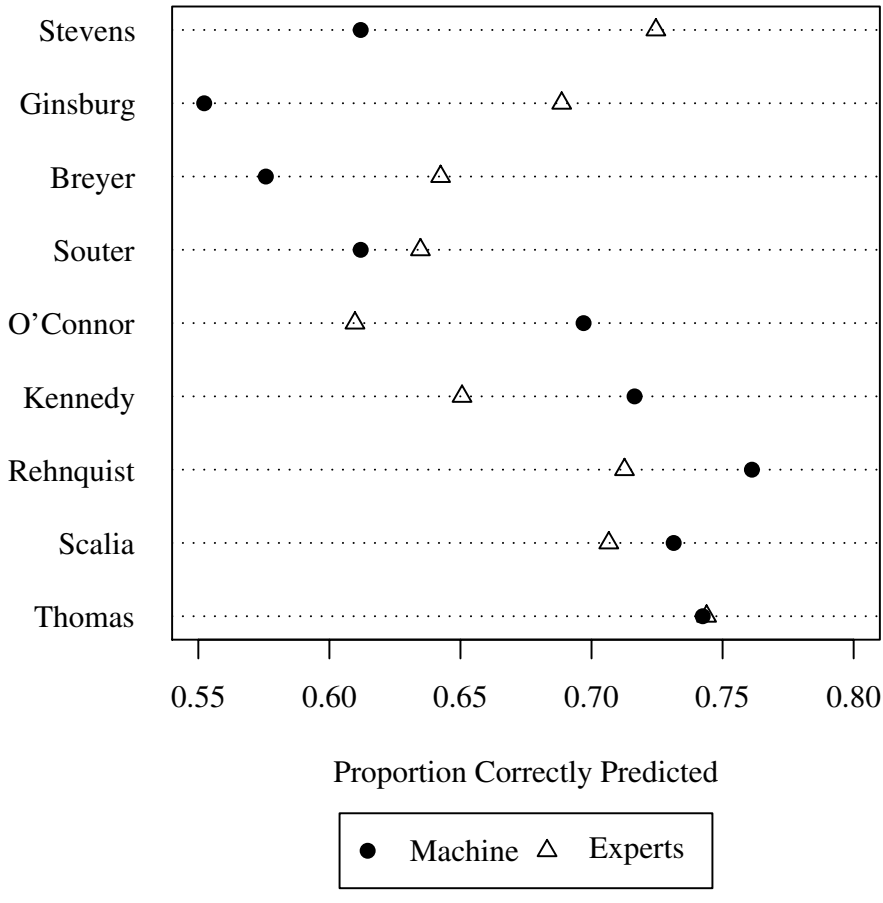
The machine and the experts varied considerably in the accuracy of their forecasts for different Justices. Figure 2 graphs the proportion of correctly predicted votes by the machine and the experts for each individual Justice. As is apparent, the experts did worst at predicting Justice O'Connor's votes among all the Justices, and considerably worse than the machine. That the legal experts found Justice O'Connor difficult to predict is not surprising—she is widely viewed as an enigmatic moderate by observers of the Court.[74] What *is* surprising is that the statistical model was able to correctly predict O'Connor's votes 70% of the time. Thus, the model seems to have captured patterns in her decisional behavior that the experts did not recognize.

Figure 2 also clearly demonstrates that the experts did better at predicting the Justices at the opposite ends of the Court's ideological spectrum. Figure 2 arrays the Justices along the vertical axis in order of increasing conservatism as estimated for the 2001 Term by Martin and Quinn.[75] The proportion of correct predictions forms a sideways V-shape, indicating that the experts were most accurate at predicting the votes of the most ideologically extreme Justices, and were least successful at forecasting the votes of the centrist Justices. Relying solely on the Justices' ideology to predict outcomes would likely produce a similar pat-

74. See, e.g., Ruth Colker & Kevin M. Scott, Dissing States?: Invalidation of State Action During the Rehnquist Era, 88 Va. L. Rev. 1301, 1345 (2002) ("Our data . . . support the commonly held view that Justice O'Connor is a moderate swing voter who cannot be described in predictable ideological terms."); Linda Greenhouse, Between Certainty & Doubt: States of Mind on the Supreme Court Today, 6 Green Bag 2d 241, 247 (2003) (describing O'Connor as "one of the Court's leading minimalists").

75. See Martin & Quinn, supra note 29. We suspect that most legal academics would generally agree with this lineup.

FIGURE 2: MACHINE AND EXPERT FORECASTS OF VOTES FOR DECIDED CASES (N=67), BY JUSTICE.



tern, suggesting that the legal experts view the Court in part in attitudinal terms. It is also possible that some other factor—perhaps some Justices' clear judicial philosophies or interpretive theories—aligns with this apparently liberal/conservative axis, making it easier for legal experts to predict the actions of the Justices on the extreme ends.[76] Figure 2 also reveals that the statistical model was much better at predicting the votes of the conservative Justices than it was with the more liberal Justices. Because the experts did much better than the machine at predicting the votes of Stevens, Ginsburg, Breyer, and Souter, the overall accuracy of the two methods across all the Justices was about the same. However, given

76. As discussed in Part IV below, the survey responses indicate that the policy preferences and judicial ideologies of the Justices were important factors in the experts' predictions, but so, too, were factors like Court precedent, statutory text, and the practical consequences of the decision.

the current composition of the Court, predicting the votes of the five conservative Justices correctly is apparently more important for getting the overall result right. An examination of the direction of the error rates reinforces this point. Both the machine and the experts over-predicted conservative outcomes, but a greater proportion of the machine's errors were in a conservative direction.[77] Despite, or perhaps because of, this conservative bias, the machine proved significantly more accurate in forecasting case outcomes.

C. *Different Issue Types, Different Results*

We also parsed our results by issue area. In doing so, we used the issue area codes assigned by Spaeth in his Supreme Court database.[78] These issue area categories may seem awkward or even arbitrary from a legal perspective, as they do not neatly track traditional doctrinal categories. Nevertheless, Spaeth's coding protocol is well-defined, and his issue area labels have been widely used by political scientists. More importantly, our statistical model utilized Spaeth's coding protocol to determine issue area codes for input into the classification trees. Thus, they provide a useful starting point for analysis.

Figures 3 and 4 display the proportion of correctly predicted case outcomes and Justice votes for issue areas with five or more cases in our sample. These figures suggest that the relative success of the two methods varies significantly depending upon the issue area. Given the small number of cases in each category, these comparisons are obviously quite sensitive to the category definitions and the coding decisions in individual cases. Nevertheless, striking deviations occurred in the judicial power and economic activity cases. The substantial variations from one issue area to another suggest that one method or the other may have a comparative advantage in predicting certain types of cases.

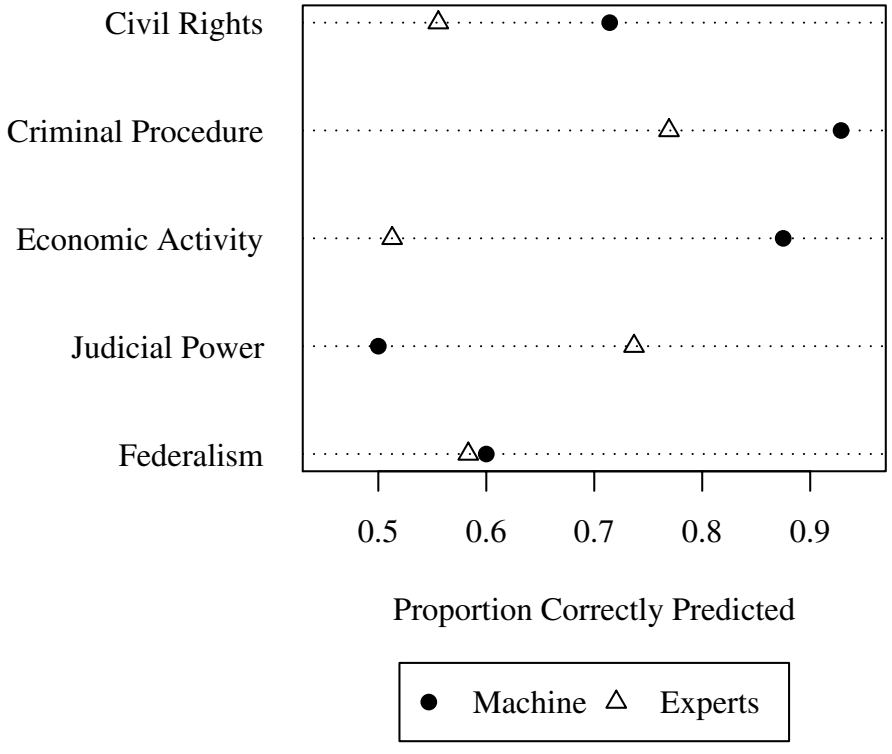
In the judicial power cases, the experts did significantly better than the machine, both in predicting case outcomes and individual votes. In these cases, the experts correctly predicted 73.7% of outcomes and 76.0% of the Justices' votes, compared with accuracy rates of 50% and 37.5% respectively for the model.

Cases in the economic activity category present the opposite picture. In this issue area, the machine's rate of correct outcome forecasts—87.5%—far exceeded that of the experts, who accurately predicted only

77. Of the cases misclassified by the machine, 18.7% were conservative outcomes that the machine had predicted would be liberal outcomes, and 81.3% were liberal outcomes that the machine had predicted would be conservative. For the experts, the figures were 33.8% and 66.2%, respectively.

78. For decades, Harold Spaeth, a leading political science scholar of the Court, has classified every Supreme Court decision by, among other things, subject matter. He utilizes some 260 categories, which are in turn grouped into thirteen major categories. By "issue area" we refer to these thirteen broad categories as captured in the variable "VALUE" in Spaeth's database. See Spaeth, Documentation, supra note 45, at 41.

FIGURE 3: MACHINE AND EXPERT FORECASTS OF CASE OUTCOMES FOR DECIDED CASES, SELECTED BY ISSUE AREA. THE ISSUE CATEGORIES ARE: CIVIL RIGHTS (*N*=14), CRIMINAL PROCEDURE (*N*=14), ECONOMIC ACTIVITY (*N*=16), JUDICIAL POWER (*N*=8), AND FEDERALISM (*N*=5).

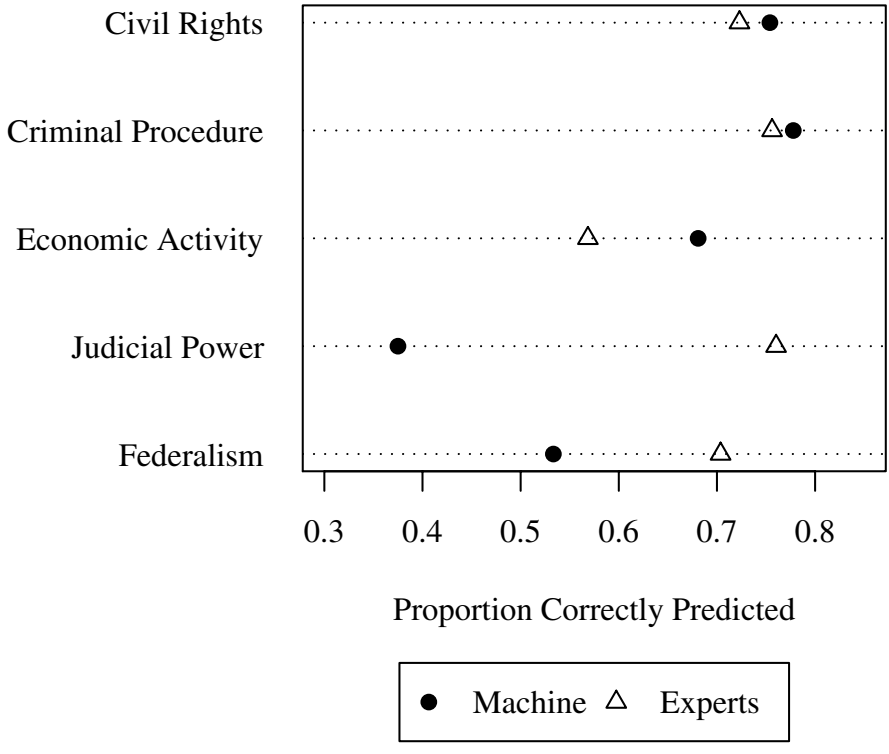


51.3% of the cases. Remarkable from a legal perspective is the widely varying subject matter of the cases encompassed within the "economic activity" issue area. The implications of the model's success across such a diverse doctrinal grouping and the experts' success in the judicial power cases is explored below in Part IV.

D. *Attorneys and Academics*

This study was designed to compare the predictive accuracy of a statistical model with a group of legal experts. In the analysis above, we treated all legal experts the same, although they have differing backgrounds and professional experiences. This group of experts included twelve specialized appellate attorneys in addition to seventy-one legal academics, and nearly half had experience clerking at the Supreme Court. These numbers are too small, and our method of case assignment within

FIGURE 4: MACHINE AND EXPERT FORECASTS OF JUDICIAL VOTES FOR DECIDED CASES, SELECTED BY ISSUE AREA. THE ISSUE CATEGORIES ARE: CIVIL RIGHTS (N=14), CRIMINAL PROCEDURE (N=14), ECONOMIC ACTIVITY (N=16), JUDICIAL POWER (N=8), AND FEDERALISM (N=5).

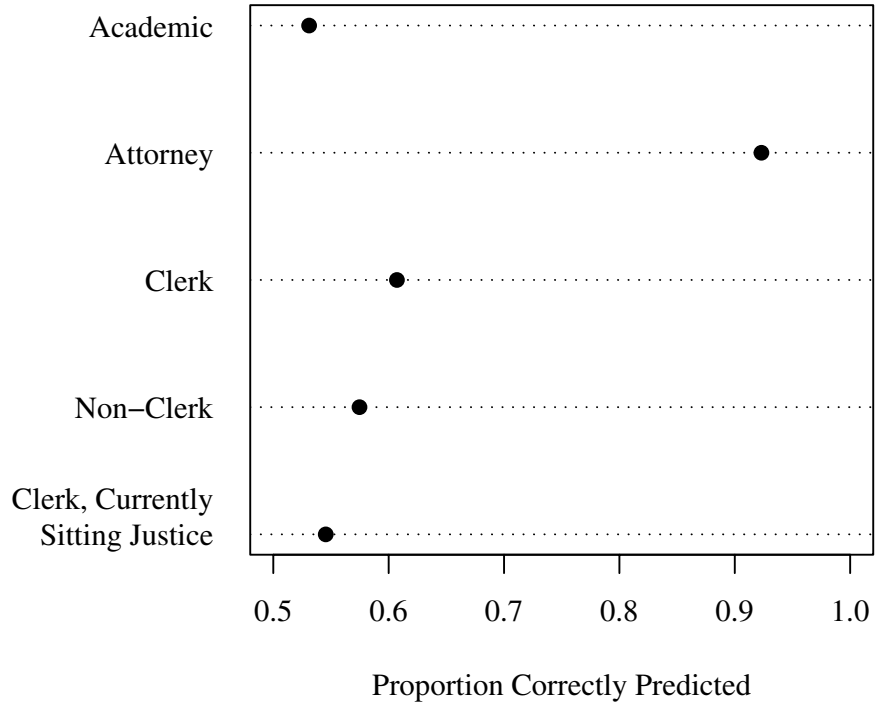


the expert pool too unsystematic,[79] to produce firm conclusions about differences between types of experts. Nonetheless, two findings are worth noting—the first for how much accuracy variation existed, the second for how little difference emerged. The small group of appellate attorneys did much better at forecasting cases than the academics, but—contrary to our hypothesis—those experts who had clerked at the Supreme Court, even fairly recently, did not demonstrate greater accuracy than the experts at large. Figure 5 displays these internal points of comparison within the expert group.

The legal academics and practicing attorneys in our pool of experts differed markedly in the accuracy of their predictions. The legal academics forecast 53% of their cases correctly, while the attorneys were correct

79. We did not distribute the relevant categories of expert—Supreme Court clerks versus non-Supreme Court clerks, academics versus attorneys—randomly across types of cases.

FIGURE 5: PROPORTION CORRECT EXPERT FORECASTS OF CASE OUTCOMES BY EXPERT BACKGROUND. THE FIGURE IS BASED ON THE FOLLOWING FORECASTS: 145 FORECASTS BY ACADEMICS; 26 BY PRACTICING ATTORNEYS; 84 BY EXPERTS WHO CLERKED FOR THE SUPREME COURT; 87 BY NON-SUPREME COURT CLERKS; AND 34 FORECASTS BY EXPERTS WHO CLERKED FOR A CURRENTLY SITTING JUSTICE.



92% of the time. This sharp difference in accuracy should be interpreted cautiously, as there were only twelve attorneys among our pool of eighty-three experts. Moreover, the process of matching experts and cases may have disproportionately assigned the attorneys to the more straightforward cases. Because of this concern, we excluded the "judicial power" cases (where the experts generally did very well) from the analysis, but the performance gap between the attorneys and the academics remained about the same.[80] Although a more systematic comparative study might not produce such a large gap, it is plausible that the two groups actually differ in their predictive accuracy. The practicing attorneys who participated in this project are appellate lawyers who appear regularly before the Supreme Court. Prediction of Supreme Court outcomes, in order to advise clients and develop litigation strategies, is an important element of their professional role. By contrast, for most legal academics, even those

80. The adjusted accuracy rate was 94.4% for the attorneys and 53.7% for the academics.

whose scholarship centers on the Supreme Court, forecasting cases is a minor component of their work—both in terms of time and importance.[81]

We similarly expected to see a difference in predictive accuracy between nonclerks and former Supreme Court clerks, or at least for those who clerked within the last ten to fifteen years for one of the currently sitting Justices. The year spent inside the Court might, we thought, confer a more nuanced understanding of the Justices' preferences and legal philosophies that might aid in prediction. Roughly half of the participating experts clerked for a Justice on the Supreme Court, and of those, twenty-one clerked for a Justice on this current Court. Our data do not show any clear difference between these groups of experts: Former Supreme Court clerks predicted 61% of cases correctly, compared with 57% for nonclerks. The subcategory of "clerk, currently sitting Justice" got 54% of case outcomes correct. Just as with the attorney/academic comparison, we did not design the project to assess this intragroup difference and so this finding is extremely tentative.

IV. Discussion and Implications

This project, which centers on the comparative prediction of Supreme Court outcomes, began with a different kind of prediction made in a faculty lounge months before the October 2002 Term began. The two law professor authors (Ruger and Kim) on this study listened to the two political science authors (Martin and Quinn) present their findings from a retrospective empirical analysis of Court decisions. The work was illuminating and rigorous, but we were skeptical about the utility of a model that left out so much legal and factual nuance in its analysis of cases. Our prediction at the time, which developed into this study, was that a sophisticated group of legal experts could forecast outcomes in specific cases more accurately than a statistical model that failed to take into account particular legal text or doctrine.

By now it is evident that our initial prediction was wrong, at least with respect to this iteration of the experiment. Although a different outcome

81. This variance in expert performance underscores one limitation in placing too much weight on the comparative nature of this study. We assembled a group with expertise in the substantive law that comes before the Supreme Court; we did not necessarily assemble a group of Supreme Court *prediction* experts. Many of the experts who are well-accomplished in analyzing the Court's work expressly disclaimed any particular predictive ability. It might be possible, with sufficient focus and enough trial and error, to assemble a different group of legal experts in future Terms who would perform as well as, or better than, a statistical model such as this one. The comparative results from the 2002 Term are worthy of notice, and perhaps reaffirm that Supreme Court prediction is no easy task, but many of the general implications we discuss in the next Part would apply even if the forecasting results of the two methods were much closer. The fact remains that a model that was purposely blinded to specific doctrine and text predicted 75% of case outcomes accurately, and this result is interesting even independent of the experts' performance.

might obtain in another round, the model's success in predicting the 2002 Term is impressive, and has forced us to reassess our thoughts about the potential benefits and implications of such a generalized model of Supreme Court voting. With our skepticism significantly dispelled, we now shift from speculating how a machine that is blinded to doctrinal, textual, and procedural particularity would do *poorly* at predicting cases, to asking why it did so *well*—from assumptions about the particulars the machine misses to curiosity about the underlying generalities it appears to have captured. What did the model recognize about the patterns of the Justices' behavior that the experts did not? And what does this tell us about how we might better observe and understand Supreme Court decisionmaking? We discuss a few more specific implications of this study first before considering these broader questions.

A. *Practical Applications and Limitations*

The Supreme Court is a critical institution in American society, and its decisions have wide ramifications on a host of social, political, and economic areas. Those who have an interest in Court outcomes—whether that interest is personal, professional, financial, or intellectual—would have interest in a machine that could do well at predicting outcomes. This notion is obvious and almost tautological: Those who would like to predict Supreme Court outcomes would have an interest in a machine that does predict them, and that interest would presumably increase with the model's accuracy rate. This last point compels an important qualifier about the model's utility for predicting actual outcomes: It does well at assessing *probable* outcomes across a diverse array of cases, but it does not achieve certainty or anything close to it—the model missed a quarter of the case outcomes in the 2002 Term.[82] Moreover, as this iteration of the experiment showed, the model's outputs themselves are potentially subject to human error manifested in programming errors or data input mistakes.[83] Accordingly, we suspect that a general predictive model would be of some use to those with specific interests in case outcomes, but would only complement, and not replace, the tools that attorneys and others currently use to assess probable results. For potential litigants, the analogy might be to the techniques of scientific jury re-

82. Considering that the Supreme Court reverses more often than it affirms, a naïve model might predict a reversal in every case. For the 2002 Term, such a model would have achieved a 72% accuracy rate, one almost as good as our statistical model. In other recent Terms, however, such a "reverse" model would have been less successful: The aggregate reversal rate for all argued cases in the ten terms preceding the October 2002 Term was 63%, and in only two of those ten terms did the reversal rate exceed 70% (1996 and 2001). The low point over the preceding decade in terms of reversal rate was the October 1993 Term, where the Court reversed in only 51% of the cases it heard argued. Reversal rates were generated from data available in Spaeth's Supreme Court database, see supra note 45.

83. See supra notes 56–58 and accompanying text for a discussion of the particular programming bug that occurred last Term.

search, which is used as background by many litigants with sufficient resources, although ultimate juror selection choices are made by attorneys.

One clear limitation of this model's general predictive power is that it corresponds to this specific group of Justices on the Supreme Court. The model succeeded to the extent that it did because it was able to discern meaningful patterns in the past voting behavior of these nine Justices that correlated reasonably well with future votes. This achievement was facilitated by the existence, rare in American history, of over 600 decisions from the same nine Justices sitting together since 1994. A change in the Court's composition would make the model-building process much more difficult. Certainly a new Justice's decision tree would differ from the retiring one. Moreover, a change in the Court's personnel would likely affect the behavior of the holdover Justices as well. In the current model some of the Justices' decision trees are expressly dependent upon the predicted votes of other sitting Justices. Not only would those trees have to be re-estimated, but the strategic environment in which each Justice votes would likely shift, such that their past behavior might no longer provide a good guide to their future behavior.

There are additional challenges to creating a successful model of this sort to predict outcomes in the federal circuit courts, where the rotating panel system might confound ready model-building. This model depends on the ability to observe voting coalitions in a large number of cases. On a court where panel composition—and therefore, the judges' strategic environment—varies, such patterns might be more difficult to capture. Moreover, the observable variables that proved useful for predicting Supreme Court outcomes in this model are themselves keyed to features of the appeals court ruling (e.g., "circuit of origin," "direction of circuit court decision"). Creation of a model to predict lower court decisions would require identification of different ex ante predictors.

In addition, a more fundamental feature of Supreme Court decisionmaking may limit the applicability of this type of predictive model to other courts. The statistical model is intentionally ignorant of the particularities of doctrine and text (note that we do not say it is ignorant of "law"—this is a different question discussed below). A predictive method that ignores specific text and doctrine might be expected to do relatively well—especially when compared with the predictions of experts trained in interpreting doctrine and text—in a decisional setting where those specific commands are relatively ambiguous. Cases before the Supreme Court are typically those that present novel factual situations or in which persuasive legal authority exists on both sides. Because the law in these cases is more ambiguous, and therefore less constraining, than at the trial or circuit court level, forecasting Supreme Court decisionmaking likely involves significantly different considerations than predicting outcomes elsewhere in the legal system. It may well turn out that taking account of specific legal arguments is more important for accurate forecasting of trial and circuit court decisions.

B. *Different Blindness, Different Vision*

This study compares two very different methods for predicting Supreme Court behavior. One method—the statistical model—is quite obviously blinded to a host of case-specific considerations that might aid prediction. Another glance at the trees in Appendix A confirms this—not only is the model oblivious to legal nuance, it also ignores the specific facts and procedural posture of the cases. The model is thus bound to miss a significant number of cases every Term where specific legal and factual idiosyncracies push the Justices outside of the normal patterns that the model captures. The judicial power cases are most likely examples of this.

As discussed in Part III, the experts substantially outperformed the model in predicting both case outcomes and votes in the judicial power cases. The cases in this category[84] generally involved technical issues of procedure in which the rule of decision was unlikely to directly implicate broad policy debates outside the legal system. In such situations, the legal experts arguably have a comparative advantage over the machine. For example, in *Breuer v. Jim's Concrete of Brevard, Inc.*,[85] all three experts correctly predicted a 9-0 affirmance, while the machine predicted a 5-4 reversal. This case raised the question of whether statutory language conferring concurrent jurisdiction in state and federal courts barred removal to federal court of an action initiated by the plaintiff in state court.

Cases like *Breuer*, more than most, likely turn on highly particularized features of the case—perhaps conventional "legal" factors such as statutory text and stare decisis—that the experts were able to recognize and incorporate into their decisionmaking process. In fact, survey responses in the judicial power cases had a markedly different profile. Experts in these cases indicated that the Justices' policy preferences and ideology played a relatively lesser role, and statutory text a greater role, in their predictions than for expert respondents in last Term's cases taken as a whole. The machine, limited to the gross features of the case, likely missed the very specific factors on which these outcomes turned. In fact, two of the experts explained their predictions in *Breuer* by pointing to several highly specific features of the case, none of which could possibly be captured by the sorts of variables utilized by the statistical model.[86]

---

84. The eight judicial power cases include Nguyen v. United States, 123 S. Ct. 2130 (2003) (consolidated with *Phan v. United States*); Beneficial National Bank v. Anderson, 123 S. Ct. 2058 (2003); Breuer v. Jim's Concrete of Brevard, Inc., 123 S. Ct. 1882 (2003); Roell v. Withrow, 123 S. Ct. 1696 (2003); Jinks v. Richland County, 123 S. Ct. 1667 (2003); Dole Food Co. v. Patrickson, 123 S. Ct. 1655 (2003); United States v. Bean, 537 U.S. 71 (2002); and Syngenta Crop Protection, Inc. v. Henson, 537 U.S. 28 (2002).

85. 123 S. Ct. 1882.

86. One expert wrote:

> The question presented seems straightforward and the opinion below seems correct. The Court granted expedited review for this case and no other out of nine cases in which it granted cert. in the same day. This could suggest that the Court regards this as a simple case to brief, argue and decide.

Despite the experts' success in the judicial power cases, the model was more accurate across a broad range of cases. Although the machine could not account for and process certain bits of specific information, there are likely countervailing limitations on individual experts' ability to assess and process all the various types of information available to them.[87] This might be manifested in at least two ways. First, experts might overemphasize analysis of specific doctrine and text which—although important—might not alone offer the best guide to predicting the close cases the Supreme Court considers. Second, when experts do look beyond legal doctrine and text to consider other factors in prediction, the limits of human cognition may make it difficult to recognize and correctly assess the broader patterns that correlate with the Justices' decisions.

On the first point, it is clear that the experts took into account specific legal considerations that the machine ignored. When asked "in making your prediction, what sources of information did you consult?" they overwhelmingly pointed to traditional legal materials, such as court decisions, statutes, and the briefs in the case.[88] When specifically asked about traditional legal factors such as precedent and statutory text, significant majorities of the expert responses rated them important factors in their decisions. For example, 69% of expert responses rated as important "Supreme Court precedent on point" in the cases in which such authority existed.[89] Similarly, in cases in which it was relevant, 54% of expert responses indicated that statutory text was an important factor in their prediction.

However, doctrine and text may be uniquely indeterminate grounds for predicting Supreme Court decisions given that institution's case selection criteria and its place in the American judicial hierarchy.[90] Most of the issues the Court hears have already been decided in contrary ways by panels of lower court judges, and there is no higher judicial authority to

---

The other expert explained:

> I am influenced by the position of the United States supporting affirmance, the clear federal interest at stake, and the absence of good reasons of policy for these cases to be left in state court at the discretion of the plaintiff. The strength of those considerations, I think, will overwhelm the predilections of some of the Justices against removal.

87. One obvious limitation is time. Given the uncompensated nature of the task and the competing demands on their time, different experts likely devoted different amounts of time and attention to their prediction efforts. It is quite possible that the amount of time and attention devoted to the task affected the accuracy of the predictions. We simply had no way of either controlling or measuring the level of effort invested by individual experts.

88. Virtually every one of the expert responses to this particular question cited these types of legal materials. Somewhat less frequently, they also reported that they had read scholarly commentary or spoken with colleagues before reaching their prediction.

89. See Appendix D.

90. To the extent that this is true, the fact that most of our participants are experts on the *law*, not the *Court*, may have contributed to the experts' relatively poorer showing overall.

ensure compliance with a particular interpretive regime.[91] The Court's cases are typically "hard" cases, for which precedent and legal text offer ambiguous or conflicting answers. It is hardly surprising in this context that doctrine and text would be unreliable cues for prediction. Karl Llewellyn stressed these limitations of doctrine in making predictions of future behavior, saying of legal scholars:

> Our own blindness is the correlative blindness of the insider. We insist, even among ourselves, on treating the cases primarily as repositories of doctrine. They are that, and of course we need both to know it and to use our skills in the refining of that ore. But opinion by opinion . . . case by case, the reports offer vastly more than data about the prevailing rules of law.[92]

Llewellyn's insight—that court decisions reflect more than just "the prevailing rules of law"—is now accepted by many in the legal academy. However, merely recognizing that other, nonlegal factors matter is insufficient to produce consistently accurate prediction. Consider how the experts applied presumptions about the Justices ideological preferences in making predictions. Judicial ideology was important for many experts relative to many case predictions. Substantial majorities—65% and 54.2% respectively—of expert responses rated the "policy preferences of the Justices" and "the conservative or liberal ideologies of the individual Justices" as important factors in their forecasts. As revealing as what the experts said in this regard is what they actually did in predicting individual Justice's votes. As discussed in Part III, the experts' accuracy rate by individual Justice was markedly higher at the ends of the Court's ideological spectrum than it was in the middle, producing the neat sideways V-shaped curve visible in Figure 2, a pattern consistent with traditional attitudinal assumptions.

That the legal experts have difficulty with Justices O'Connor and Kennedy is hardly surprising. Some Justices have articulated clear interpretive philosophies that give strong cues about their votes even in close cases. The moderate Justices, however, often appear to observers to rely on narrower, idiosyncratic, and case-specific rationales. One leading legal scholar has characterized the center of the current Court as "minimalist," maintaining that the Justices at "the analytical heart of the current Court [ ] have adopted no 'theory' of constitutional interpretation."[93] Justice O'Connor presents particular problems in this regard—her central role on the current Court is widely regarded as important but also as

---

91. This is not to say specific law is irrelevant or unimportant, merely that it is often ambiguous.

92. Llewellyn, Common Law Tradition, supra note 26, at 355–56.

93. See Cass R. Sunstein, The Supreme Court: 1995 Term—Foreword: Leaving Things Undecided, 110 Harv. L. Rev. 4, 14 (1996) (emphasis omitted) (enumerating O'Connor, Kennedy, Souter, Breyer, and Ginsberg as the minimalist Justices); see also Cass R. Sunstein, One Case at a Time: Judicial Minimalism on the Supreme Court 8–10 (1999) (explaining that minimalism seeks to avoid "broad rules and abstract theories," instead going only as far as "necessary to resolve a particular dispute").

enigmatic and unpredictable by many observers.[94] The prediction results suggest that the experts relied on highly general attitudinal assumptions to supplement their assessment of legal factors, but such blunt attitudinal assumptions are of limited utility in predicting the center of the Court. What is needed is a more systematic and nuanced recognition of the voting patterns of the moderate Justices, and this is difficult for human experts to discern from case-by-case analysis.

This point applies more broadly to factors beyond merely law or ideology. The experts' decisionmaking processes are most accurately characterized as heterogeneous and multi-factorial. The experts relied on a variety of different factors, to differing degrees across experts and cases. Seven of the listed factors were rated as important in a majority of expert responses.[95] However, a great deal of individual variation existed in how various factors were weighted. Almost all of the listed factors—twenty-three of twenty-six—were rated "very important" by one or more experts,[96] and their handwritten comments reported additional factors that influenced their predictions, such as the unique facts of a case, observed trends in a particular area of the law, or the Court's overall reversal rate. Moreover, the same expert, asked to predict outcomes in different cases, weighted the various factors differently, sometimes rating a factor such as precedent or ideology as "not at all important" in one case and "very important" in the next.[97] Thus, rather than utilizing a uniform set of decision criteria, the experts appeared to take into account a large number of factors, giving them varying weight depending upon the particular facts of each case.

We initially thought that this ability to consider multiple factors in individualized ways would help the experts' performance, but that appar-

---

94. See Colker & Scott, supra note 74, at 1345 (noting that O'Connor is "a moderate swing voter who cannot be described in predictable ideological terms"); Mark A. Graber, Rethinking Equal Protection in Dark Times, 4 U. Pa. J. Const. L. 314, 328 (2002) (describing "[t]he minimalism of Justice O'Connor and, to a lesser extent, Justice Kennedy"); Merrill, supra note 38, at 629 n.228 (observing that O'Connor is "likely to be the median voter in contested cases").

95. These seven factors are listed in bold in the table summarizing the survey results. See Appendix D. Of the seven, some are clearly legal and others reflect attitudinal assumptions, but two factors are not easily classified. The "interpretive theories of the Justices" could be viewed merely as heuristics that help them reach their desired policy outcomes, or, alternatively, as philosophies adopted for reasons internal to the law that potentially constrain the Justices from pursuing naked preferences. Similarly, the "practical consequences of the decision" seems to encompass both the real world policy implications of a decision, as well as more limited effects confined to the legal system itself.

Interestingly, some factors that the experts generally did not believe to be important—such as the preferences of Congress or the Executive Branch—were identified in the literature as affecting the Court's strategic environment. See Appendix D.

96. See id.

97. The varying weights reflect the experts' judgments about *which* factors matter for that particular case. For example, one expert, correctly predicting a 9-0 reversal in Massaro v. United States, 123 S. Ct. 1690 (2003), wrote, "I think this is a case where practicalities, as reflected in prior U.S. position, will trump ideological predispositions."

ently did not happen across the board. Rather than conferring an advantage, perhaps the experts' ability to consider highly particularized information interfered with their predictive success. Considerable research in cognitive psychology has demonstrated the limits of human cognition.[98] People often make poorer rather than better decisions when confronted with more information, because they may shift to simpler, less accurate decision strategies, or may become distracted by less relevant information.[99] Experts are also vulnerable to these effects.[100] Moreover, like all humans, experts are beset by various biases—such as availability biases or confirmation biases—that affect their judgments.[101] The use of heuristics, though adaptive over the long run, may lead to poor judgments in particular cases. Especially in situations like this—involving large amounts of information and multiple relevant factors—cognitive limits may hamper the experts' ability to systematically analyze and account for the impact of multiple relevant factors.

C. *Finding "Reckonability" Without Reference to Doctrine and Text*

The experts' close attention to legal doctrine turned out to be insufficient to predict reliably the Court's decisions. For all the insight to be gained from careful reading of cases, such attention to the details of doctrine and text may blind legal experts to broader patterns in the cases which are visible only at a higher level of generality. Similarly, resort to simple attitudinal assumptions will help predict the votes of some Justices but not others, and not those who matter most for outcomes. Despite the degree of discretion afforded the Supreme Court, and despite the Court's often confounding ideological equipoise on many issues, the statistical model succeeded in recognizing patterns in the Justices' behavior

---

98. A great deal of recent legal scholarship discusses this cognitive psychology literature and its implications for the law. See, e.g., Chris Guthrie, Framing Frivolous Litigation: A Psychological Theory, 67 U. Chi. L. Rev. 163 (2000); Chris Guthrie, Jeffrey J. Rachlinski & Andrew J. Wistrich, Inside the Judicial Mind, 86 Cornell L. Rev. 777 (2001); Christine Jolls, Cass R. Sunstein & Richard Thaler, A Behavioral Approach to Law and Economics, 50 Stan. L. Rev. 1471 (1998); Russell B. Korobkin, Behavioral Analysis and Legal Form: Rules vs. Standards Revisited, 79 Or. L. Rev. 23 (2000); Russell B. Korobkin & Thomas S. Ulen, Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics, 88 Cal. L. Rev. 1051 (2000); Jeffrey J. Rachlinski, Heuristics and Biases in the Courts: Ignorance or Adaptation?, 79 Or. L. Rev. 61 (2000); Symposium, Empirical Legal Realism: A New Social Scientific Assessment of Law and Human Behavior, 97 Nw. U. L. Rev. 1075 (2003).

99. See Troy A. Paredes, Blinded by the Light: Information Overload and Its Consequences for Securities Regulation, 81 Wash. U. L.Q. 417, 437–43 (2003) (summarizing social science research on information overload).

100. Id. at 453–58 (citing research indicating that while experts may be better at selectively filtering information than lay people, they can become overloaded and in fact make worse decisions than lay people in certain circumstances).

101. Although we did not generate the data necessary to explore these theories, one can speculate that the legal experts' tendency to focus on recent, salient cases in a particular area of the law, and their normative commitments to certain outcomes as more desirable, might bias their judgments of what the Court is likely to do.

sufficient to predict correctly the outcomes of 75% of the cases. Thus, the machine—to a greater extent than the legal experts—appears to have captured a measure of Llewellyn's elusive "reckonability."[102] This result suggests that accurate prediction depends on the identification of factors that correlate with the Justices' decisions at an intermediate level of generality: less specific than "the statute in this case says x" and less general than "Justice Y is a conservative." And on this score the model's approach to prediction worked well, particularly so for the important Justices at the center of the Court. How it might have done so merits further exploration.

The model had one clear advantage in discerning these patterns with respect to the current Rehnquist Court: the hundreds of past cases in which the Justices' voting behavior was revealed. But data collection is only the first step; accurate prediction requires the selection of variables that correlate sufficiently with behavior so that they can be used to forecast unknown future cases. A workable model requires that these variables be few in number. The model succeeded to the extent that it did because it identified case characteristics—observable before decision—that correlate with outcomes across a broad variety of cases.

As discussed above, the final model relied on only six variables: circuit of origin, identity of the petitioner, identity of the respondent, ideological direction of the decision below, claim of unconstitutionality, and issue area. To the legal eye, these six variables are an odd set of factors on which to base predictions about the Court's decisions. Most of the variables seem overly blunt and bereft of any analysis of doctrinal or textual specificity. But although the model's analysis is *more* general than a particularistic legal perspective, it is significantly *less* general than the basic attitudinal assumption that some Justices are more liberal and some are more conservative. Instead the model relies heavily on variables of intermediate generality.

For one thing, the model disaggregated the Justices and considered behavior patterns independently rather than as a linear ideological array, as attitudinal studies do expressly and the legal experts appear to have done implicitly here. Unlike the traditional attitudinal model, our statistical model did not rigidly adhere to the assumption that the Justices are arrayed linearly along some ideological space. Rather, each Justice's classification tree was estimated separately, and the trees differ dramatically from one another, both in their shape and content. Consistent with the attitudinal model, the statistical model includes a variable for the liberal or conservative orientation of the decision below in order to capture how the Justices' ideological preferences influence their willingness to reverse the outcome. Once again, however, the statistical model's classification trees capture the influence of ideology in a more subtle way than simply predicting that conservative Justices will seek conservative outcomes and

102. See supra note 26 and accompanying text.

vice versa. Other variables, such as the identity of the parties and the circuit of origin, interact with the basic liberal or conservative nature of the decision, allowing the Justices' differing preferences to lead to different responses depending upon the type of litigant or origin of the case.[103]

To see how this approach might have been successful, consider two areas in which the model did particularly well: predicting the critical votes of Justices O'Connor and Kennedy, and predicting outcomes in the doctrinally heterogeneous category of "economic activity" cases. The model's success in predicting case outcomes was due in large part to its accuracy in predicting the votes of Justices O'Connor and Kennedy. The model got Justice O'Connor's vote right 70% of the time and Justice Kennedy's 72%, as compared with accuracy rates of 61% and 65%, respectively, for the experts. For the Court's centrist Justices, neither close analysis of legal authority nor simple ideology offer much predictive power. Instead, using the six general case characteristics, the statistical model appears to have captured patterns in their voting behavior.

Consider Justice O'Connor's classification tree.[104] The first decision point is blunt—it predicts a vote to reverse whenever the lower court decision is "liberal." But as to "conservative" opinions under review, the model's classification of Justice O'Connor's vote is both nuanced and systematic. For instance, the model predicts that Justice O'Connor's vote is likely to differ depending upon the circuit of origin. If a conservative decision comes from the Second, Third, District of Columbia, or Federal Circuit, she is likely to affirm. If it arises from one of the other circuits, she is more likely to reverse. This does not imply that O'Connor votes to affirm *because* a case is from the Second Circuit, but only that her votes tend to correlate with the origin of the case in this way. "Circuit of origin," then, works as a proxy for some aspect of the legal process—not directly observable—that influences outcomes. One interpretation, consistent with attitudinal explanations, is that judges in the Second, Third, District of Columbia, and Federal Circuits are more closely aligned with Justice O'Connor's moderate-conservative ideology. Alternatively, "circuit of origin" may capture some other differences—perhaps variations in legal culture, the types of rationales offered by judges, or deference given to particular appellate judges—that are more consistent with Justice O'Connor's legal philosophy.

---

103. The legal experts for the most part disregarded these variables relied on by the machine. These three—circuit of origin, identity of the petitioner, and identity of the respondent—were deemed unimportant in the large majority of expert responses. Respectively, 64%, 69.1%, and 75.5% of expert responses rated circuit of origin, identity of petitioner, and identity of respondent a one or two on a five-point Likert scale (oneNºt at all important; five=very important). See Appendix D. To the extent that the interaction of these variables also captures the Justices' preferences, the experts, by largely ignoring them, appear to have incorporated attitudinal assumptions into their predictions in a less nuanced way than the statistical model.

104. See supra Figure 1.

Another area of remarkable success for the model was its accuracy in predicting the set of cases within the broad category of "economic activity" as coded by Spaeth.[105] The "economic activity" category is "largely commercial and business related; it includes tort actions and employee actions vis-à-vis employers."[106] Spaeth's categories often seem peculiar to legal academics, precisely because they do not take account of what seem to be obvious legal distinctions—for instance, two cases arising under two different federal statutes might be clumped together without reference to glaring differences in the statutory texts.

The sixteen "economic activity" cases, as coded by Spaeth, illustrate this divergence. Viewed from a legal perspective, the cases are highly disparate and offer few commonalities for use in analysis or prediction. The category includes cases ranging from *Eldred v. Ashcroft*,[107] addressing the constitutionality of the Copyright Term Extension Act, to *State Farm Mutual Automobile Insurance Co. v. Campbell*,[108] challenging the constitutionality of punitive damages as excessive under the Due Process Clause, to *Yellow Transportation v. Michigan*,[109] involving interpretation of the Intermodal Surface Transportation Efficiency Act. Other cases in this issue area turned on questions of bankruptcy law,[110] the Federal Trademark Dilution Act,[111] interpretation of an arbitration agreement,[112] and the False Claims Act,[113] among others.[114] For lawyers, the cases appear to involve a broad array of seemingly unrelated statutory and doctrinal issues.

As discussed above, the machine did *much* better than the experts at predicting the outcomes in the sixteen cases in this subject area—87.5% correct to the experts' 51.3%. Particularly impressive was the model's remarkable success in predicting the votes of the three most important Justices on the current Court: Chief Justice Rehnquist and Justices O'Connor and Kennedy. For those three jurists at the center-right of the

105. See Spaeth, Documentation, supra note 45, at 40–41.

106. Id. at 42.

107. 537 U.S. 186 (2003).

108. 123 S. Ct. 1513 (2003).

109. 537 U.S. 36 (2002).

110. Archer v. Warner, 123 S. Ct. 1462 (2003); FCC v. Nextwave Pers. Communications, Inc., 537 U.S. 293 (2003).

111. Moseley v. V Secret Catalogue, Inc., 537 U.S. 418 (2003).

112. PacifiCare Health Sys., Inc. v. Book, 123 S. Ct. 1531 (2003); Howsam v. Dean Witter Reynolds, Inc., 537 U.S. 79 (2002).

113. Cook County v. United States, ex rel Chandler, 123 S. Ct. 1239 (2003).

114. Fitzgerald v. Racing Ass'n of Cent. Iowa, 123 S. Ct. 2156 (2003) (state tax code/Equal Protection Clause); Hillside Dairy v. Lyons, 123 S. Ct. 2142 (2003) (state milk pricing regulations/Commerce Clause); Dastar Corp. v. Twentieth Century Fox Film Corp., 123 S. Ct. 2041 (2003) (Lanham Act); Black & Decker Disability Plan v. Nord, 123 S. Ct. 1965 (2003) (Employee Retirement Income Security Act); Pharm. Research Mfrs. of Am. v. Walsh, 123 S. Ct. 1855 (2003) (Medicaid); Norfolk & W. Ry. Co. v. Ayers, 123 S. Ct. 1210 (2003) (Federal Employers' Liability Act); Pierce County v. Guillen, 537 U.S. 129 (2003) (highway safety/Commerce Clause).

current Court, the model's success rates in predicting their votes in the economic activity cases were 86.7%, 75%, and 81.2%, respectively, compared with the experts' accuracy rates of 55.6%, 51.3%, and 51.3%, respectively. In this doctrinally disparate area, the machine's method appears to have captured some commonality among the cases that is overlooked by more narrowly defined legal categories. Spaeth has described his own goal in creating such a typology as capturing "the subject matter of the controversy rather than its legal basis. . . . The objective is to categorize the case from a public policy standpoint, a perspective that the legal basis for decision . . . commonly disregards."[115] The fact that the machine recognized such clear patterns in some of the Justices' votes in the economic activity cases suggests that there is some analytical gain to grouping them together, despite their lack of connection from a textual or doctrinal perspective. The general grouping captures something relevant to prediction that a more highly specified legal classification scheme misses.

D. *What Does Prediction Say About "The Nature of Law"?*

In the ways explored above, a study such as this one offers some lessons about predicting cases, and perhaps also more generally about methods of observing and studying the Court. Much less clear is whether predictive exercises have anything to say about the nature of law itself. Frederick Schauer suggested that they might in a theoretical essay a few years ago. He maintained that "by looking at the various ways in which a person might seek to predict the future behavior of judges, we will have discovered something important about the type and size of the chunks with which law makes its decisions, and, less directly, something equally important about the nature of law itself."[116] We share much of his belief that by comparing means of prediction, we can assess the "type and size of the chunks with which law makes its decisions" and discern broader patterns of judicial decisionmaking. The results discussed above suggest that—at least for the Supreme Court—bigger, more general "chunks" may produce better predictions than attention to specific doctrine and text.

However, anything our study has to say about the "nature of law itself" is highly indirect. Part of the model's success lay in its ability to identify observable factors that *correlate* with decisions—it was indifferent to underlying theories of causation or judicial motivation. This indifference may actually aid prediction, because many of the possible causal factors (such as stare decisis and judicial ideology) that might influence judicial decisions in particular cases are extremely difficult to observe and measure directly. Precisely measuring the manner in which doctrine, text, judicial ideology, institutional setting, and other factors interact to

115. Spaeth, Documentation, supra note 45, at 41.

116. Schauer, supra note 23, at 774–75.

influence decisionmaking is probably impossible, but the model has done the next best thing by identifying easily observable features that correlate with decisionmaking at a reasonably high accuracy rate. That such correlations exist and can be measured does not mean that other more obscure causal factors are not in fact driving the Court's decisions.

That said, the correlated factors are neither random nor irrelevant, nor are they unrelated to actual causation. At the very least, for instance, the fact that Justice O'Connor's votes appear to vary consistently with circuit of origin across a variety of cases suggests that there are some underlying differences among the circuits that warrant further exploration. That these factors are correlated with actual behavior lends some credence (but in no way is a critical test of) spatial theories of voting. Indeed, the fact that circuit of origin, issue area of the case, etc., are related to the *types* of cases heard by the Justices is not surprising because the Justices choose the cases they hear. While we could only hypothesize about why, for example, the Court takes certain types of cases from certain circuits, the results of this study suggest some avenues to explore in empirically modeling the agenda process.

Moreover, under any theoretical conception that regards law as consisting at least in part of what judges do, proxies that reliably predict what they *will* do in the future are worth considering as baselines or guideposts of "law," whether or not we can imagine them as "law" themselves. We noted above that Holmes's famous *Path of the Law* address stressed the importance of prediction without offering much to advance the project of prediction. But in a much earlier statement discussing what constitutes the basis of "law," Holmes maintained that "[a]ny motive for [judicial] action . . . which can be relied upon as likely in the generality of cases to prevail, is worthy of consideration as one of the sources of law."[117] We have amended Holmes's insight—in the ellipsis above he enumerated four traditional "legal" sources ("constitution, statute, custom, or precedent")—but, thus updated, his point seems highly applicable to a study such as this one. This study has not sought to determine ultimate causation for, or full explanation of, what the Supreme Court does. But it has identified several broad factors that "can be relied upon as likely in the generality of cases to prevail." The model's proxies are not "law" themselves, but they may capture something close to it that is interesting and illuminating for those who study the Court.

In this way a reliable general prediction method makes some indirect contribution to our sense of what "the law" might be—or at least the law of the Supreme Court. But of course this is an incomplete picture of both the law and the Court, and we note here a few things that our study emphatically does *not* say. The first limitation applies to description and analysis. Prediction of outcomes alone gives little insight into the content

117. Oliver Wendell Holmes, Jr., Book Notice: The Law Magazine and Review, 6 Am. L. Rev. 723, 724 (1872) (review of Frederick Pollock).

of the Court's opinions, and such content matters greatly for application in the lower courts and for a ruling's reception in broader American society. This study focused on binary results to the exclusion of opinion content, and so offers no insights regarding this important aspect of law. Moreover, the focus on outcomes alone placed this contest solidly on the statistical model's turf. Legal experts, particularly academics, spend their time analyzing what the courts say, not merely what they do. In this project, we asked them to focus solely on what the Court might do, reducing their predictions to a simple "affirm" or "reverse" forecast.[118] Comparing the machine and the legal experts solely on the basis of their vote and outcome predictions is perhaps unfair to the experts, because it privileges a certain type of performance and overlooks insights that, though valuable, are more difficult to capture quantitatively.[119] More fundamentally, beyond binary outcomes the study has nothing to tell us about how the Court is likely to shape, explain, and justify its important decisions.

Just as the study's findings are limited even in this complete descriptive sense, they are also largely bereft of specific normative import. The study focused on prediction of what the Court would do, not what it should do. Much of the best work in legal scholarship is expressly normative, offering the academy, the bar, and the Justices themselves persuasive visions of how the law might, and should, look. Calls for a return to greater scholarly normativity are occasionally heard in political science as well. Our study speaks only indirectly to such normative scholarship. Perhaps, by outlining certain broad patterns in the Justices' past behavior, studies such as this one will provide an additional point of background data for those who would more comprehensively assess and critique the Court's jurisprudence. The significant uncertainty in both prediction methods is relevant here, since no Justice appears wholly predictable, and they may depart from prior patterns in ways that we applaud or criticize.

A different kind of normative disclaimer is also necessary. In focusing on predicting the Court's decisions, we do not mean to suggest that predictability itself is the paramount goal either for those within the judicial process or those who study it. Some degree of regularity or predictability in judicial decisionmaking is important to the functioning of the system and the ability of people to anticipate the consequences of their

---

118. Although there were reasons for using this format, see supra Part II.A.3, some experts were clearly and understandably frustrated by this limitation. One expert wrote, "You cannot possibly do this without breaking down the questions. Here, the Court is likely to split on the questions . . . that won't fit a neat yes/no model."

119. For example, legal experts correctly predicted that Ford Motor Co. v. McCauley, 537 U.S. 1 (2002), would be dismissed as improvidently granted, that Justice O'Connor would not participate in Howsam v. Dean Witter Reynolds, Inc., 537 U.S. 79 (2002), and that Borden Ranch Partnership v. United States Army Corps of Eng'rs, 537 U.S. 99 (2002), would be affirmed by an equally divided Court after Justice Kennedy recused himself. Because we could not easily classify these outcomes in our coding scheme, the legal experts got no "credit" for anticipating these developments.

actions. Such concerns underlie the value placed on stare decisis and the rule of law. However, given the crucial role of the Supreme Court in our society, absolute predictability in its decisions should not be expected, nor would it be desirable. To suggest, as some have, that a computer model, once perfected, might someday substitute for actual judges,[120] entirely ignores what makes the Supreme Court a uniquely important institution. Its role in American society is not merely to process important disputes expeditiously. Rather, the ways in which it addresses those disputes—not merely through outcomes, but through its rationales, its analytical framework, and its language—both gives voice to certain values and influences public understanding of these issues. Though their interpretations are often vigorously contested, the Justices' words frame the terms of the debate. How impoverished the work of the Court would be if, for example, Justice Kennedy's sweeping opinion in *Lawrence v. Texas*[121] had been reduced to the words "we reverse." We disclaim any implication that a statistical model—however accurate—could in any way substitute for the important work that the Justices do.

Similarly, we reject the notion that the model's predictive accuracy renders irrelevant factual nuances or skillful legal argument in particular cases. The model did have some success at using general case characteristics to predict outcomes, but it also missed a full quarter of the decisions. After all, its predictions are based on observed patterns in the Justices' behavior, not rigid certainties. There will always be some cases that depart from the general pattern, and the skillful advocate will be able to exploit distinctive facts, or make novel connections between legal principles in ways that reinforce or counteract these general trends, to the client's advantage. Moreover, a statistical model such as this one, which assumes that past behavior best predicts future behavior, is a necessarily limited way of modeling the judgments of real human beings who are capable of evolving over time. What the model can do is to assess systematically what good lawyers know already in a rough sense: that some cases are better shots than others, and relatedly, that on particular cases some Justices' votes will be easier to get than others.

Conclusion

In the manner suggested above, we think that the results of this comparative study provide interesting additive insights into the manner in which those who follow and study the Supreme Court might conceptualize its decisionmaking. The model's success here suggests that there is

120. For a notable early proposal of this sort, see Harold D. Lasswell, Current Studies of the Decision Process: Automation Versus Creativity, 8 W. Pol. Q. 381, 398 (1955) ("When machines are more perfect [than human decisionmakers] a bench of judicial robots . . . can be constructed."). Lasswell proposed building models to predict Supreme Court decisionmaking and noted that "a robot facsimile of the less repetitive members of the Court would provide a genuine challenge to the engineers." Id.

121. 123 S. Ct. 2472 (2003).

some value to assessing the Court's behavior in accordance with factors of intermediate generality—more general than particularized doctrine, text, or facts, and more specific than simple ideological assumptions. The model has discovered a few factors of such intermediate generality that track reasonably well with Supreme Court decisionmaking, and there may be others of equal or greater significance.

Beyond these possible substantive lessons, we hoped through this explicitly interdisciplinary study design to create a project that would be of interest to the two groups of scholars who study the Supreme Court most closely, and thereby to enhance the gradually increasing dialogue between our two disciplines. In a previous cycle of interdisciplinary interest, a participant in a *Harvard Law Review* symposium on Social Science Approaches to the Judicial Process asserted that "[i]n practice, a field seems to progress as—and if—it moves from theory to empirical data and back to theory."[122] We share this sentiment, and have acted on it, but with one amendment: The empiricism that informs theory about judicial decisionmaking is most useful if it incorporates prospective experimentation along with more common retrospective analysis.

122. Samuel Krislov, Theoretical Attempts at Predicting Judicial Behavior, 79 Harv. L. Rev. 1573, 1573 (1966).

Appendix A
Estimated Classification Trees

Figure 6: Estimated classification tree for unanimous liberal cases.

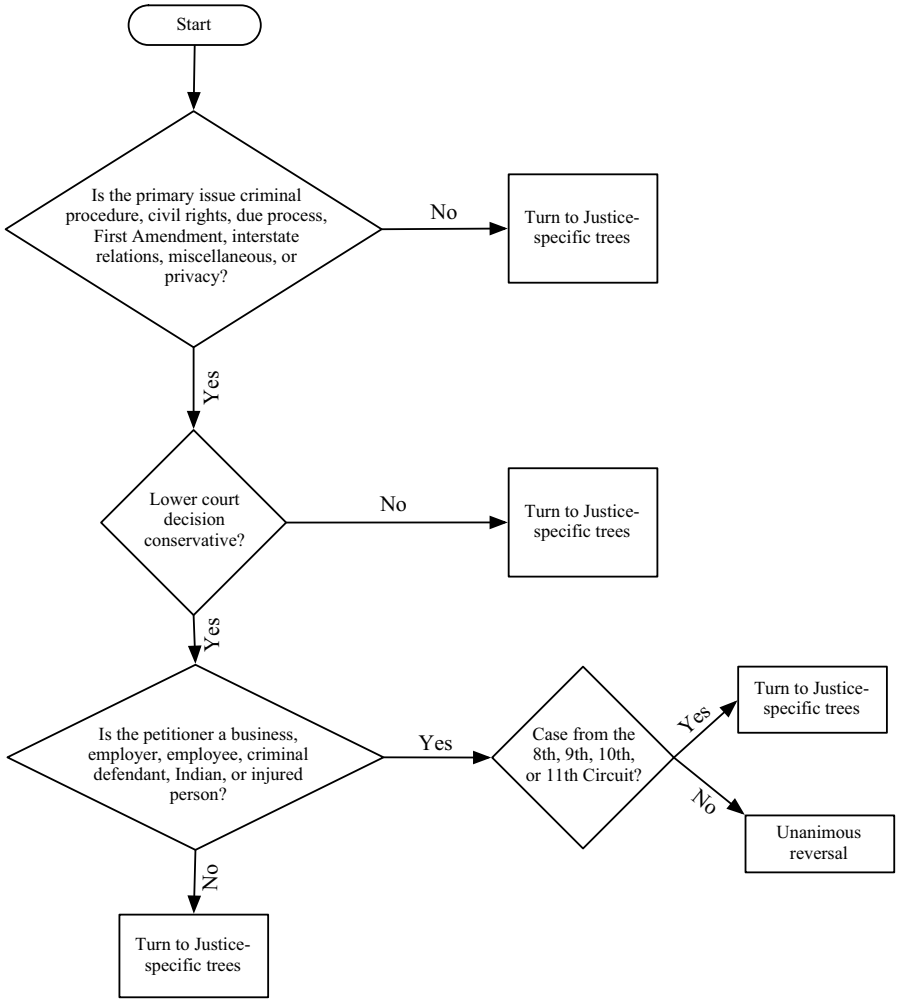


FIGURE 7: ESTIMATED CLASSIFICATION TREE FOR UNANIMOUS CONSERVATIVE CASES.

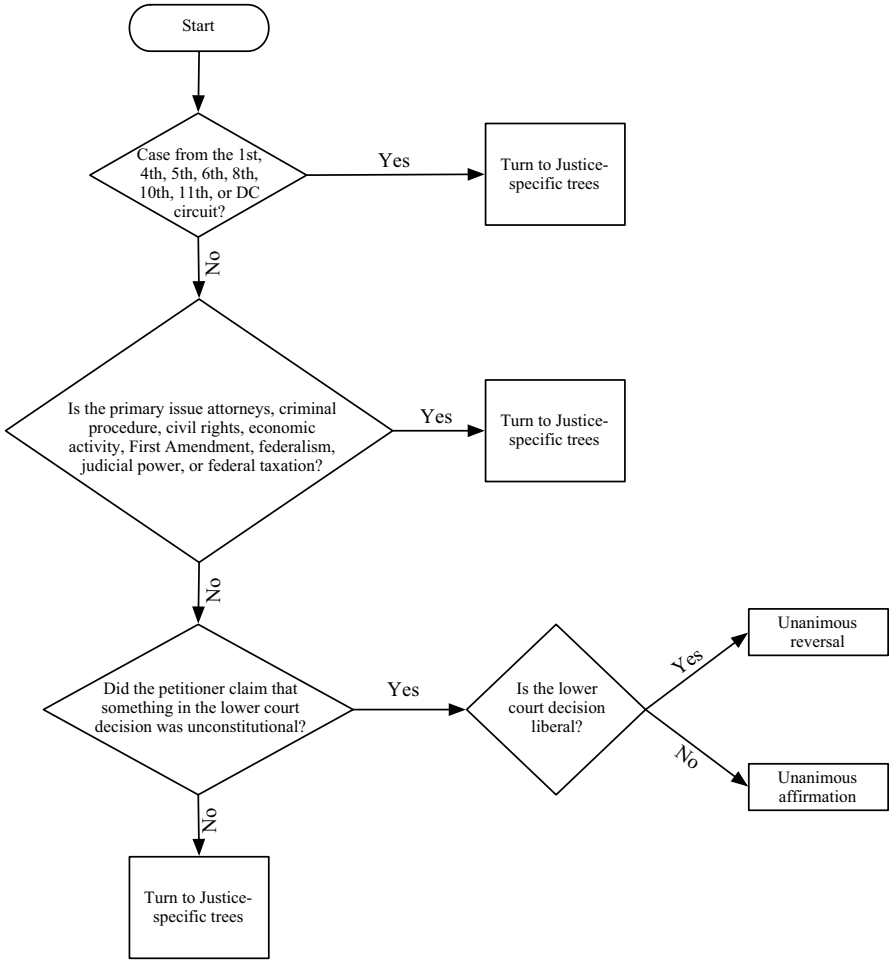


FIGURE 8: ESTIMATED CLASSIFICATION TREE FOR JUSTICE SCALIA FOR FORECASTED NON-UNANIMOUS CASES.

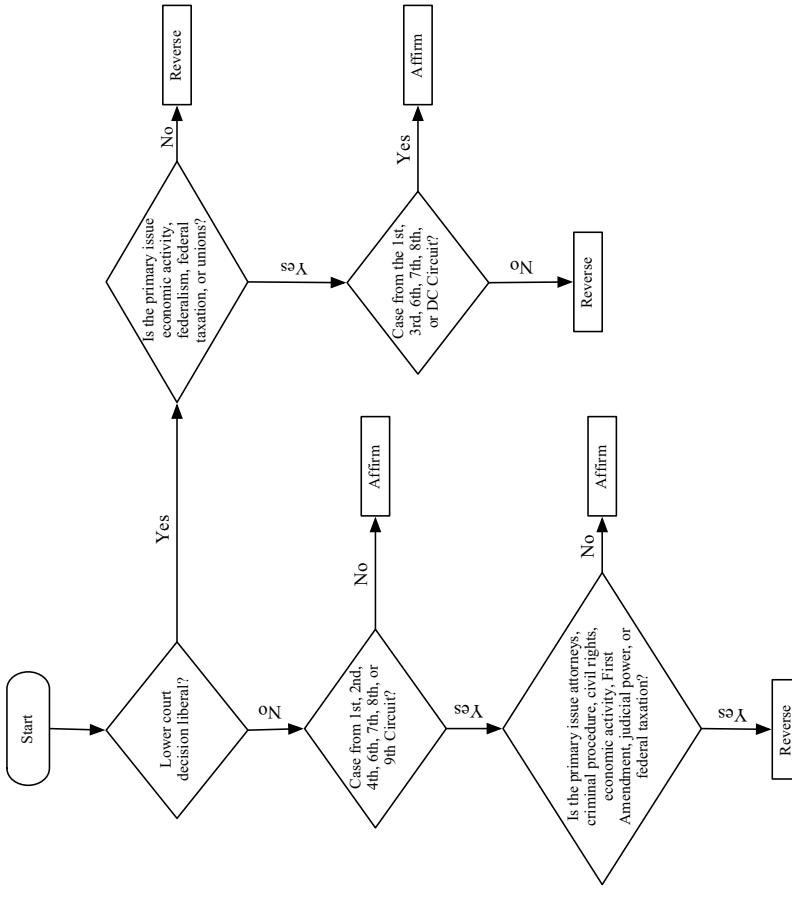


FIGURE 9: ESTIMATED CLASSIFICATION TREE FOR JUSTICE THOMAS FOR FORECASTED NON-UNANIMOUS CASES.

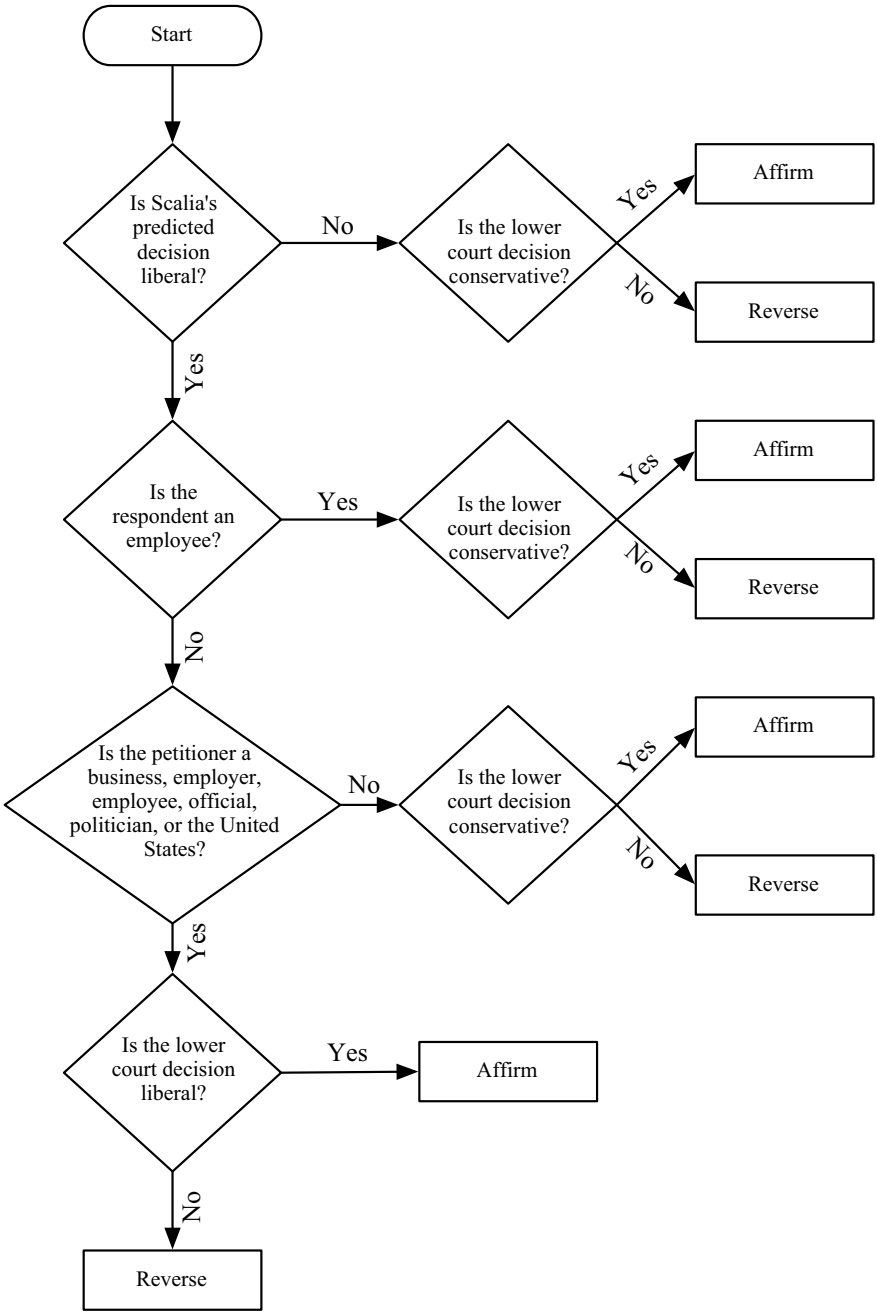


FIGURE 10: ESTIMATED CLASSIFICATION TREE FOR CHIEF JUSTICE REHNQUIST FOR FORECASTED NON-UNANIMOUS CASES.

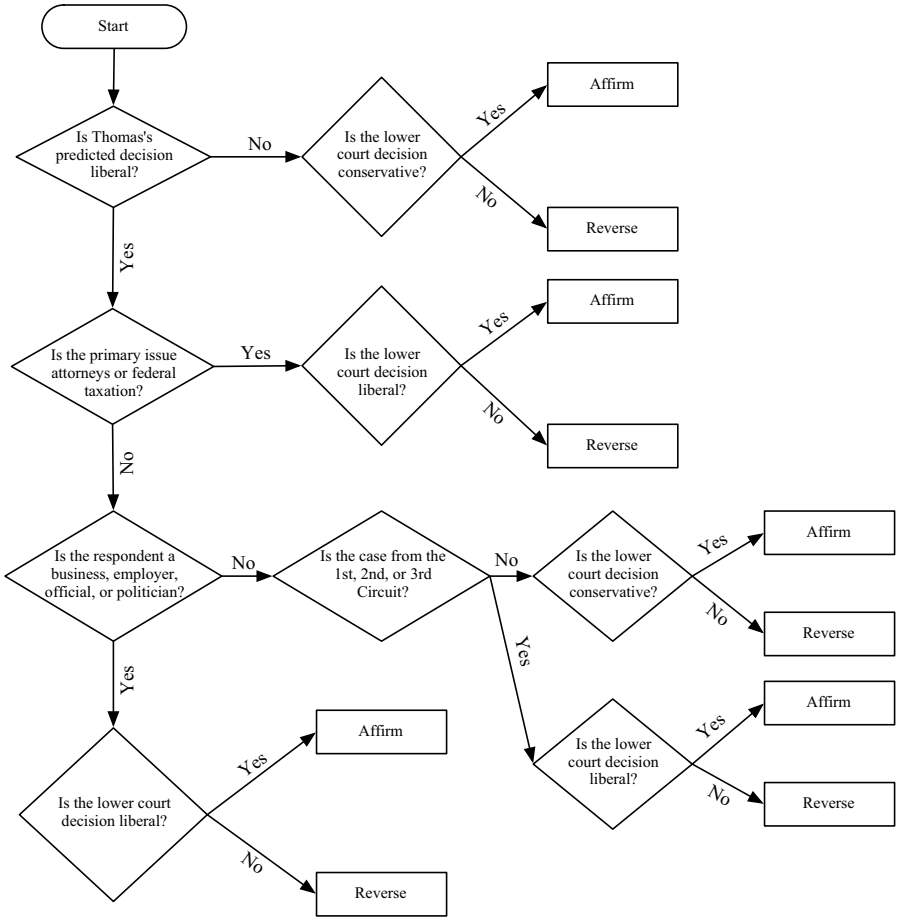


FIGURE 11: ESTIMATED CLASSIFICATION TREE FOR JUSTICE STEVENS FOR FORECASTED NON-UNANIMOUS CASES.

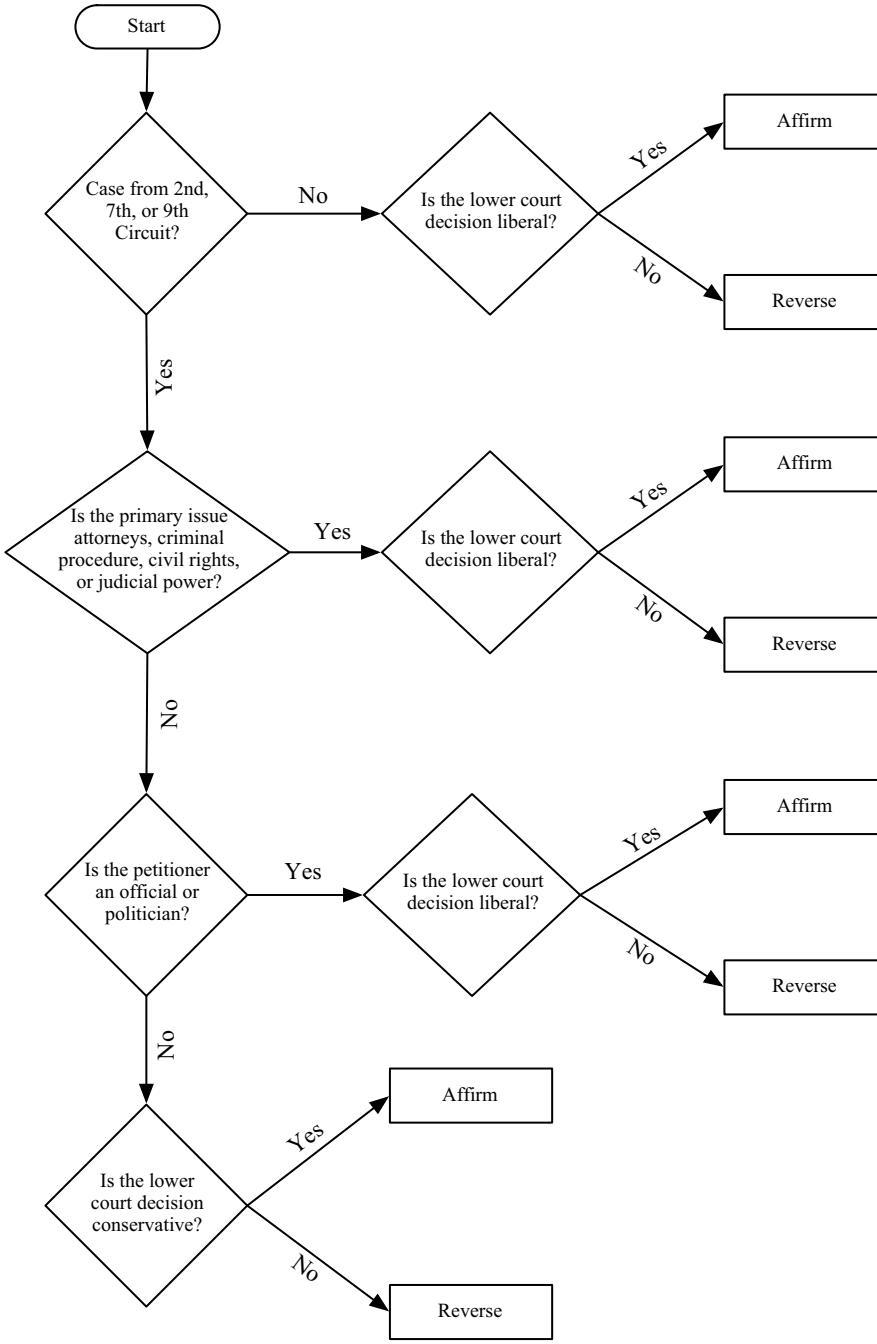


FIGURE 12: ESTIMATED CLASSIFICATION TREE FOR JUSTICE O'CONNOR FOR FORECASTED NON-UNANIMOUS CASES.

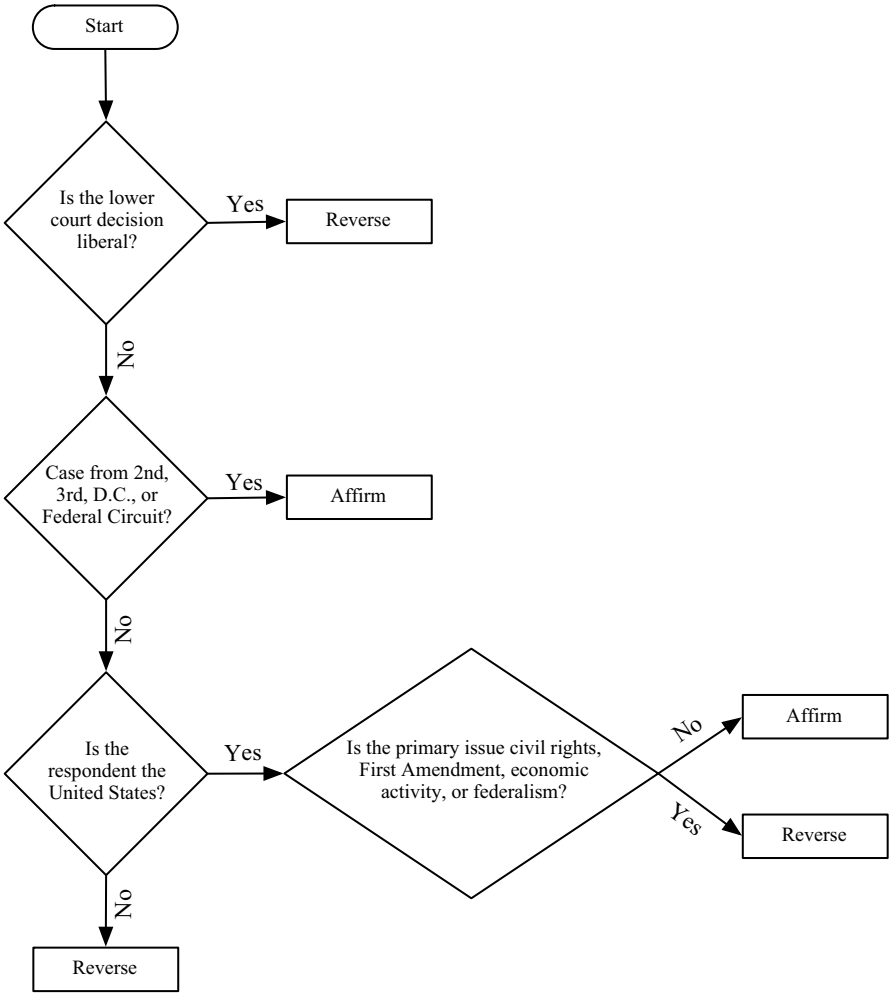


FIGURE 13: ESTIMATED CLASSIFICATION TREE FOR JUSTICE GINSBURG FOR FORECASTED NON-UNANIMOUS CASES.

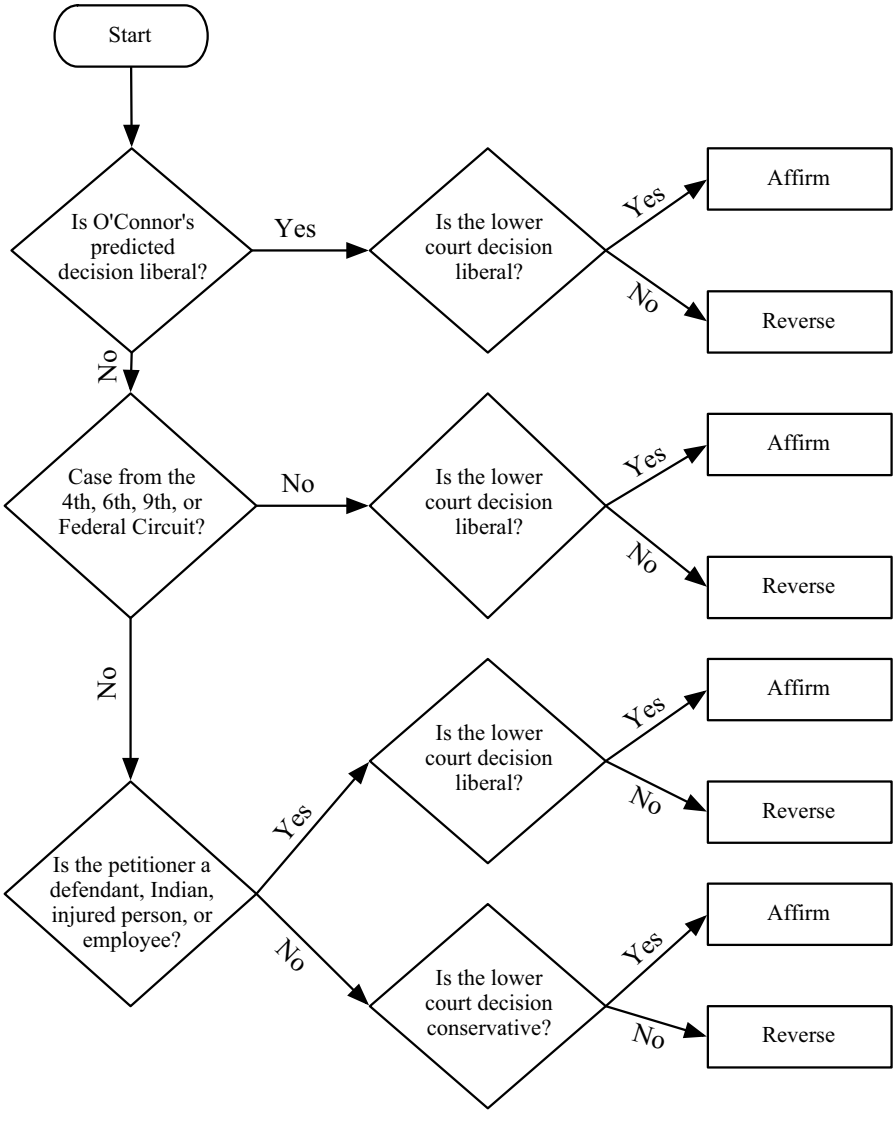
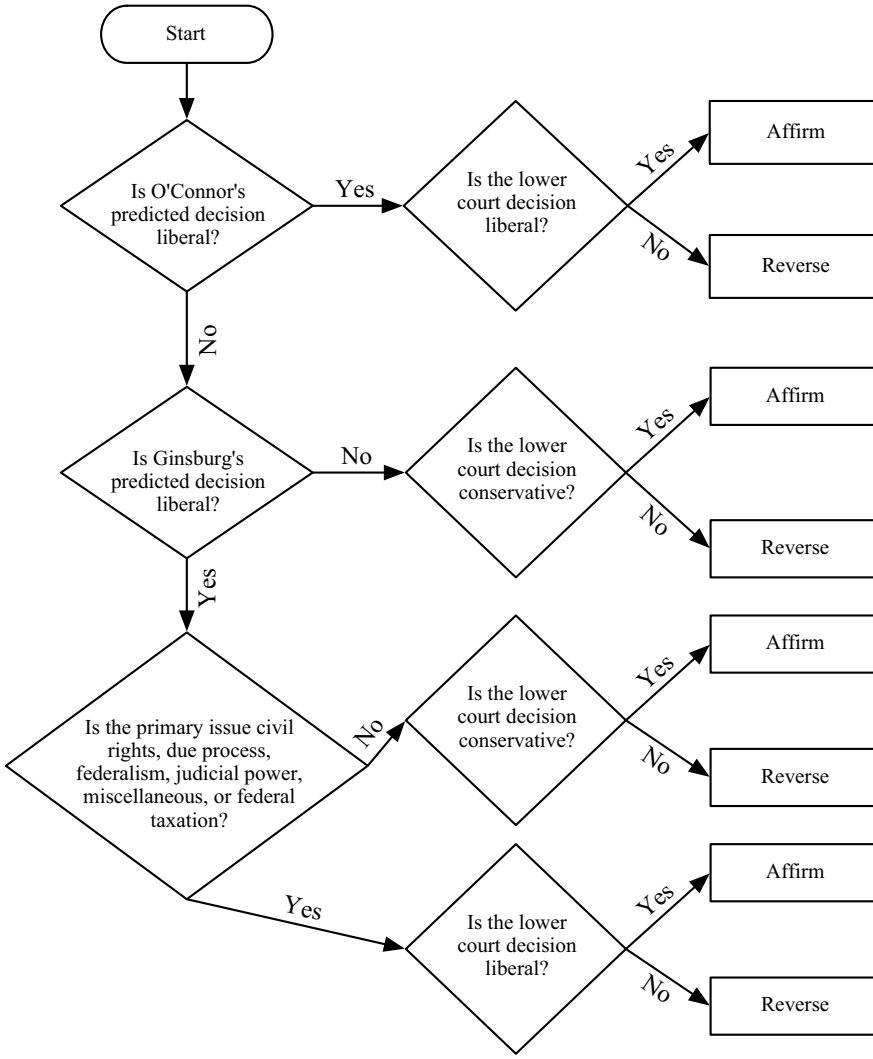


FIGURE 14: ESTIMATED CLASSIFICATION TREE FOR JUSTICE BREYER FOR FORECASTED NON-UNANIMOUS CASES.



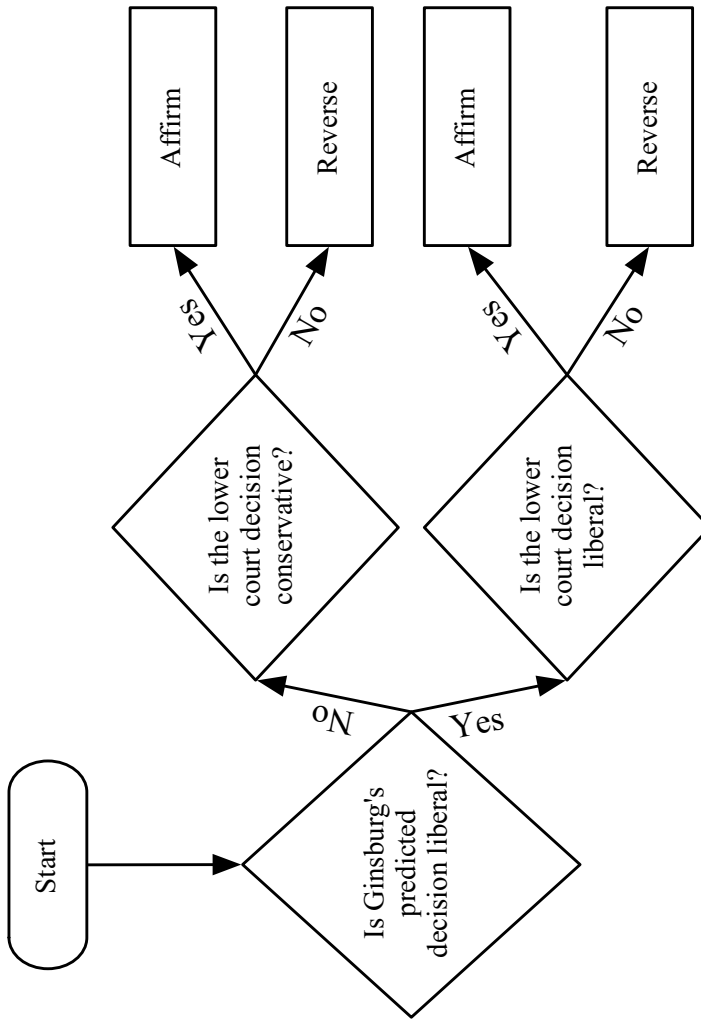
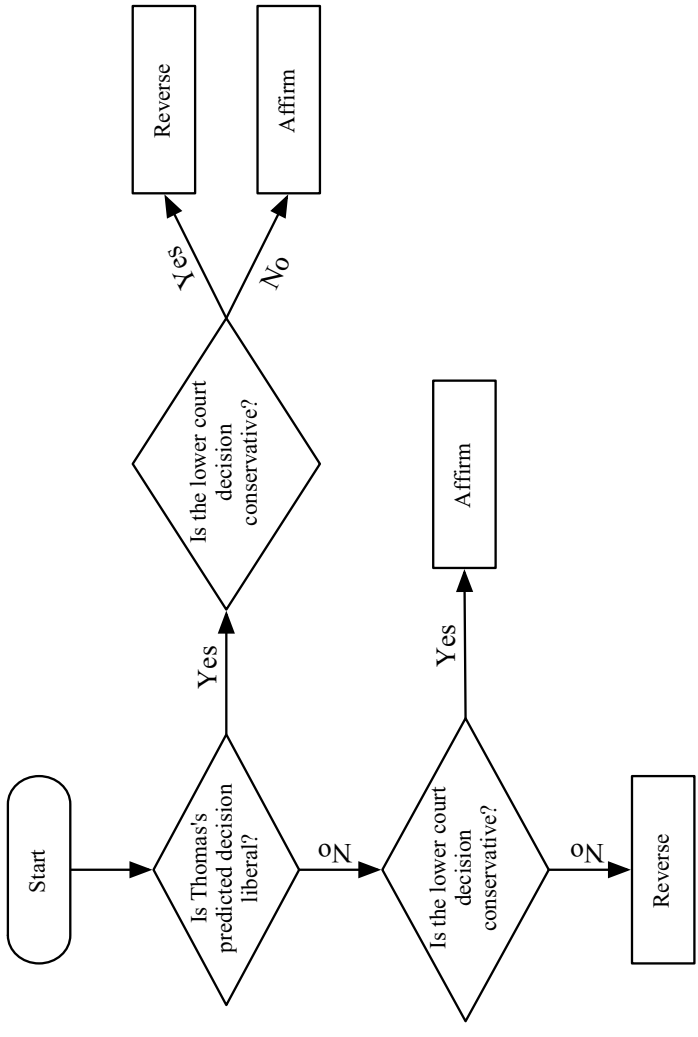


FIGURE 15: ESTIMATED CLASSIFICATION TREE FOR JUSTICE SOUTER FOR FORECASTED NON-UNANIMOUS CASES.

FIGURE 16: ESTIMATED CLASSIFICATION TREE FOR JUSICE KENNEDY FOR FORECASTED NON-UNANIMOUS CASES.



APPENDIX B
Legal Expert Participants*

Rachel E. Barkow, New York University School of Law
David J. Barron, Harvard Law School
Anthony J. Bellia Jr., University of Notre Dame Law School
Yochai Benkler, Yale Law School
James F. Bennett, Bryan Cave LLP, Saint Louis, Missouri
Paul Schiff Berman, University of Connecticut School of Law
Stephanos Bibas, University of Iowa College of Law
John H. Blume, Habeas Assistance and Training Project / Cornell Law School
Mary Ann Bobinski, University of Houston Law Center
Beth S. Brinkmann, Morrison & Foerster LLP, Washington, D.C.
Rebecca L. Brown, Vanderbilt University School of Law
Daniel J. Capra, Fordham Law School
Erwin Chemerinsky, University of Southern California Law School
Jesse H. Choper, University of California at Berkeley School of Law
Thomas Colby, George Washington University Law School
David D. Cole, Georgetown University Law Center
Brannon P. Denning, Cumberland School of Law
Neal E. Devins, William & Mary School of Law
Laura Dickinson, University of Connecticut School of Law
Michael C. Dorf, Columbia Law School
Christopher R. Drahozal, University of Kansas School of Law
Rochelle Cooper Dreyfuss, New York University School of Law
Theodore Eisenberg, Cornell Law School
William N. Eskridge, Jr., Yale Law School
Katherine Hunt Federle, Ohio State University Michael E. Moritz College of Law
Alan L. Feld, Boston University School of Law
Jonathan S. Franklin, Hogan & Hartson LLP, Washington, D.C.
Philip P. Frickey, University of California at Berkeley School of Law
Charles Fried, Harvard Law School
Kenneth S. Geller, Mayer, Brown, Rowe & Maw, Washington, D.C.
Heather K. Gerken, Harvard Law School
David H. Getches, University of Colorado School of Law
John C. P. Goldberg, Vanderbilt University School of Law
Roger L. Goldman, Saint Louis University School of Law
Thomas C. Goldstein, Goldstein & Howe, Washington, D.C.
David J. Gottlieb, University of Kansas School of Law
Margaret M. Harding, Syracuse University College of Law
Pamela Harris, O'Melveny & Myers LLP, Washington, D.C.
Melissa Hart, University of Colorado School of Law
Neal K. Katyal, Georgetown University Law Center
Jay P. Kesan, University of Illinois College of Law
Nancy J. King, Vanderbilt University School of Law
Sylvia A. Law, New York University School of Law
Robert M. Lawless, University of Nevada, Las Vegas School of Law
Douglas Laycock, University of Texas School of Law
Richard J. Lazarus, Georgetown University Law Center

---

* Expert affiliations listed are as of the date of publication.

James S. Liebman, Columbia Law School
Arnold H. Loewy, University of North Carolina School of Law
Deborah C. Malamud, New York University School of Law
Jeremy Maltby, O'Melveny & Myers LLP, Los Angeles, CA
Paul Marcus, William & Mary School of Law
Stephen R. McAllister, University of Kansas School of Law
Robert P. Merges, University of California at Berkeley School of Law
Gillian E. Metzger, Columbia Law School
Geoffrey P. Miller, New York University School of Law
Paul Mogin, Williams & Connolly LLP, Washington, D.C.
Dana Muir, University of Michigan Business School
Gerald L. Neuman, Columbia Law School
Spencer Overton, George Washington University Law School
Robert V. Percival, University of Maryland Law School
Richard H. Pildes, New York University School of Law
Robert C. Post, Yale Law School
Robert K. Rasmussen, Vanderbilt University School of Law
Alan Scott Rau, University of Texas School of Law
Larry E. Ribstein, University of Illinois College of Law
Daniel B. Rodriguez, University of San Diego School of Law
Peter J. Rubin, Georgetown University Law Center
Stewart J. Schwab, Cornell Law School
Anthony J. Sebok, Brooklyn Law School
Daniel N. Shaviro, New York University School of Law
Suzanna Sherry, Vanderbilt University School of Law
Alexander Tallchief Skibine, University of Utah College of Law
Joan E. Steinman, Chicago-Kent College of Law
Charles Jordan Tabb, University of Illinois College of Law
George C. Thomas III, Rutgers School of Law—Newark
Joseph P. Tomain, University of Cincinnati College of Law
Alan Untereiner, Robbins, Russell, Englert, Orseck & Untereiner LLP, Washington, D.C.
Robert R. M. Verchick, University of Missouri—Kansas City School of Law
Eugene Volokh, University of California, Los Angeles School of Law
Robert N. Weiner, Arnold & Porter, Washington, D.C.
Robert Weisberg, Stanford Law School
Brian Wolfman, Public Citizen Litigation Group, Washington, D.C.
Barbara Bennett Woodhouse, University of Florida College of Law

Appendix C
Predictions and Outcomes in Selected Major Cases from 2002 Term

	Stevens	Ginsburg	Breyer	Souter	O'Connor	Kennedy	Rehnquist	Scalia	Thomas	Result
Lawrence v. Texas										
Model	Reverse	Reverse	Reverse	Reverse	Reverse	Affirm	Affirm	Affirm	Affirm	5-4 to Reverse
Expert 1	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Affirm	Affirm	Affirm	6-3 to Reverse
Expert 2	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Affirm	Affirm	Affirm	6-3 to Reverse
Expert 3	Reverse	Reverse	Reverse	Reverse	Affirm	Affirm	Affirm	Affirm	Affirm	5-4 to Affirm
Actual	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Affirm	Affirm	Affirm	6-3 to Reverse
Grutter v. Bollinger										
Model	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	5-4 to Reverse
Expert 1	Affirm	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	5-4 to Affirm
Expert 2	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	5-4 to Reverse
Expert 3	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	5-4 to Reverse
Actual	Affirm	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	5-4 to Affirm
Gratz v. Bollinger										
Model	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	5-4 to Reverse
Expert 1	Reverse	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	6-3 to Reverse
Expert 2	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	6-3 to Reverse
Expert 3	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	5-4 to Reverse
Actual	Affirm	Affirm	Reverse	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	6-3 to Reverse
Ewing v. California										
Model	Reverse	Reverse	Reverse	Reverse	Reverse	Affirm	Affirm	Reverse	Affirm	6-3 to Reverse
Expert 1	Reverse	Reverse	Affirm	Reverse	Affirm	Affirm	Affirm	Affirm	Affirm	6-3 to Affirm
Expert 2	Reverse	Reverse	Reverse	Reverse	Affirm	Affirm	Affirm	Affirm	Affirm	5-4 to Affirm
Expert 3	Reverse	Reverse	Reverse	Reverse	Affirm	Affirm	Affirm	Affirm	Affirm	5-4 to Affirm
Actual	Reverse	Reverse	Reverse	Reverse	Affirm	Affirm	Affirm	Affirm	Affirm	5-4 to Affirm
Lockyer v. Andrade										
Model	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	8-1 to Reverse
Expert 1	Affirm	Affirm	Affirm	Affirm	Reverse	Affirm	Reverse	Reverse	Reverse	5-4 to Affirm
Expert 2	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	5-4 to Reverse
Expert 3	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	7-2 to Reverse
Actual	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	5-4 to Reverse
Sell v. United States										
Model	Reverse	Reverse	Reverse	Reverse	Affirm	Affirm	Affirm	Affirm	Affirm	5-4 to Affirm
Expert 1	Reverse	Reverse	Reverse	Reverse	Affirm	Reverse	Affirm	Affirm	Affirm	5-4 to Reverse
Expert 2	Reverse	Affirm	Affirm	Reverse	Affirm	Reverse	Affirm	Affirm	Affirm	6-3 to Affirm
Actual	Reverse	Reverse	Reverse	Reverse	Affirm	Reverse	Reverse	Affirm	Affirm	6-3 to Reverse
Connecticut v. Doe										
Model	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	9-0 to Reverse
Expert 1	Affirm	Affirm	Reverse	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	6-3 to Reverse
Expert 2	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	5-4 to Reverse
Expert 3	Affirm	Reverse	Reverse	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	7-2 to Reverse
Actual	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	9-0 to Reverse
Smith v. Doe										
Model	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	8-1 to Reverse
Expert 1	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	5-4 to Reverse
Expert 2	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	5-4 to Reverse
Expert 3	Affirm	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Affirm	Affirm	7-2 to Affirm
Actual	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	6-3 to Reverse
Stogner v. California										
Model	Reverse	Reverse	Reverse	Reverse	Reverse	Affirm	Affirm	Reverse	Affirm	6-3 to Reverse
Expert 1	Reverse	Reverse	Reverse	Reverse	Affirm	Affirm	Affirm	Affirm	Affirm	5-4 to Affirm
Expert 2	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	9-0 to Reverse
Expert 3	Reverse	Affirm	Reverse	Reverse	Affirm	Affirm	Affirm	Reverse	Reverse	5-4 to Reverse
Actual	Reverse	Reverse	Reverse	Reverse	Reverse	Affirm	Affirm	Affirm	Affirm	5-4 to Reverse
Scheidler v. NOW										
Model	Affirm	Affirm	Reverse	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	6-3 to Reverse
Expert 1	Affirm	Affirm	Affirm	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	7-2 to Affirm
Expert 2	Affirm	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	5-4 to Affirm
Actual	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	8-1 to Reverse
Nevada Dep't of HR v. Hibbs										
Model	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	9-0 to Reverse
Expert 1	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	5-4 to Reverse
Expert 2	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	5-4 to Reverse
Expert 3	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	5-4 to Reverse
Actual	Affirm	Affirm	Affirm	Affirm	Affirm	Reverse	Affirm	Reverse	Reverse	6-3 to Affirm
Demore v. Kim										
Model	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	9-0 to Reverse
Expert 1	Affirm	Affirm	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	6-3 to Affirm
Expert 2	Affirm	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	5-4 to Affirm
Expert 3	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	5-4 to Reverse
Actual	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	5-4 to Reverse
State Farm v. Campbell										
Model	Affirm	Affirm	Affirm	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	5-4 to Reverse
Expert 1	Reverse	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	Affirm	Affirm	6-3 to Reverse
Expert 2	Reverse	Affirm	Reverse	Reverse	Reverse	Reverse	Affirm	Affirm	Affirm	5-4 to Reverse
Expert 3	Reverse	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	Reverse	8-1 to Reverse
Actual	Reverse	Affirm	Reverse	Reverse	Reverse	Reverse	Reverse	Affirm	Affirm	6-3 to Reverse
Eldred v. Ashcroft										
Model	Reverse	Reverse	Affirm	Reverse	Affirm	Affirm	Affirm	Affirm	Affirm	6-3 to Affirm
Expert 1	Affirm	Reverse	Affirm	Affirm	Reverse	Reverse	Affirm	Reverse	Reverse	5-4 to Reverse
Expert 2	Reverse	Affirm	Affirm	Reverse	Reverse	Reverse	Affirm	Reverse	Reverse	6-3 to Reverse
Actual	Reverse	Affirm	Reverse	Affirm	Affirm	Affirm	Affirm	Affirm	Affirm	7-2 to Affirm

Appendix D
Summary Statistics of Survey Responses

Factor (Factors in bold were rated as important by a majority of respondents)	**Mean Response**	**% Rating Factor As Not Important (1 or 2)**	**% Rating Factor As Important (4 or 5)**
Identity of the court whose decision the Supreme Court is reviewing	2.1729	64.7	17.3
Existence of a divided court below	1.9792	66.7	10.4
Extent of disagreement in the circuits and/or state courts on the issue	2.5263	49.7	21.1
Identity of the petitioner	1.9470	71.9	13.7
Identity of the respondent	1.8788	75.8	12.1
Identity of counsel representing the parties	1.3806	92.6	4.5
Quality of the parties' briefs	1.9776	66.4	10.4
Supreme Court precedent on point	**3.8966**	**15.6**	**69.0**
Supreme Court dicta on point	**3.3947**	**23.6**	**54.4**
Other statements by the Justices in prior opinions	3.2061	31.3	49.6
Text of relevant constitutional provision(s)	2.2771	62.7	21.6
Text of relevant statute	**3.5495**	**22.5**	**54.0**
Text of relevant regulation	2.6250	52.1	35.4
Non-textual evidence of meaning of constitutional, statutory, or administrative provision (e.g., legislative history, long-standing practice, etc.)	3.1327	30.1	44.3
Interpretive theories of the Justices	**3.6364**	**19.0**	**62.9**
Practical consequences of the decision	**3.9254**	**9.7**	**73.8**
Policy preferences of the Justices on the specific issue presented	**3.6045**	**23.8**	**65.0**
The conservative or liberal ideologies of the individual Justices	**3.3282**	**29.0**	**54.2**
Public opinion on the issue	1.7967	78.1	9.8
Composition and preferences of Congress	1.3588	90.0	2.3
Composition and preferences of the Executive Branch	1.5420	84.8	4.6
The professional backgrounds of the Justices	1.6343	80.5	4.5
The personal backgrounds of the Justices	1.5970	82.1	5.2
Number of amici participating in the case	1.3984	87.8	2.4
Identity of amici participating in the case	1.7097	76.6	7.3
Position of the Solicitor General in an amicus filing	2.5204	52.0	27.6