TABLE OF CONTENTS

SYMPOSIUM

EMPirical Measures of Judicial Performance

EMPirical Measures of Judicial Performance:
An Introduction to the Symposium .... Steven G. Gey & Jim Rossi 1001

FORESEEING GREATNESS?
MEASURABLE PERFORMANCE
CRITERIA AND THE SELECTION
OF SUPREME COURT JUSTICES ......................... James J. Brudney 1015

JUDGING THE
TOURNAMENT ......................... Hon. Jay S. Bybee & Thomas J. Miles 1055

WHICH JUDGES WRITE
THEIR OPINIONS
(AND SHOULD WE CARE)? ............. Stephen J. Choi & G. Mitu Gulati 1077

EMPirical Measures of
Judicial Performance:
Thoughts on Choi and Gulati's
TOURNAMENT OF JUDGES ........................ Brannon P. Denning 1123

THE ROLE OF QUALIFICATIONS
IN THE CONFIRMATION
OF NOMINEES TO THE
U.S. SUPREME COURT ................................. Lee Epstein et al. 1145

SUPREME COURT SELECTION
AND MEASURES OF PAST
JUDICIAL PERFORMANCE .................. Daniel A. Farber 1175

JUDICIAL SELECTION BY THE NUMBERS ............. Michael J. Gerhardt 1197
FEDERAL JUDGES AND THE HEISMAN TROPHY ............. Steven Goldberg 1237

WHO JUDGES THE JUDGES? ............................................. John V. Orth 1245

JUDICIAL BEHAVIOR AND PERFORMANCE:
AN ECONOMIC APPROACH ......................... Hon. Richard A. Posner 1259

PULLING FROM THE RANKS?:
REMARKS ON THE PROPOSED USE OF
AN OBJECTIVE JUDICIAL RANKING
SYSTEM TO GUIDE THE SUPREME
COURT APPOINTMENT PROCESS ................. Hon. Bruce M. Selya 1281

DO JUDGES BEHAVE AS HOMO ECONOMICUS, AND IF SO, CAN WE
MEASURE THEIR PERFORMANCE?
AN ANTIPODEAN PERSPECTIVE
ON A TOURNAMENT OF JUDGES ....................... Russell Smyth 1299

JUDICIAL STRATIFICATION
AND THE REPUTATIONS
OF THE UNITED STATES
COURTS OF APPEALS ............................... Michael E. Solimine 1331

A TOURNAMENT OF VIRTUE ......................... Lawrence B. Solum 1365

INFORMATION AND THE SELECTION
OF JUDGES: A COMMENT ON
“A TOURNAMENT OF JUDGES” ...................... Ahmed E. Taha 1401

KEEPING SCORE: THE UTILITY OF
EMPirical MEASUREMENTS
IN JUDICIAL SELECTION ............................ David C. Vladeck 1415

INDEX TO VOLUME 32 ................................................. 1443
VOLUME 32 STUDENT AWARDS ............................. 1456
Florida State University
College of Law

The Florida State University College of Law is fully approved by the American Bar Association and is a member of the Association of American Law Schools. The College offers a three-year course of study leading to the Juris Doctor degree with certificate programs in international and environmental law. Furthermore, a Master of Laws degree (LL.M.) in American Law for Foreign Lawyers is available for foreign graduate students who have legal training.

The College of Law has approximately 700 students. Its law library contains more than 431,000 volumes and 5,150 current subscriptions.

Educational opportunities are plentiful. The location of the College—only blocks from the Capitol, the Supreme Court of Florida, a number of federal and state courts, and various administrative agencies—makes extensive externship programs possible. Students publish the Law Review, the Journal of Land Use & Environmental Law, and the Journal of Transnational Law & Policy. Moot Court and Mock Trial teams have won national honors. Each summer, the College conducts a nationally recognized Program in Law at Oxford University, England.

These resources support the goal Dean Mason Ladd identified when the College was founded in 1966: “Its one aim is excellence in the preparation and education of students so that they will acquire the knowledge, comprehension, and judgment essential to the performance of legal services in a complex and rapidly changing society.”
FLORIDA STATE UNIVERSITY COLLEGE OF LAW

ADMINISTRATION
R.E. LEMON, B.A., M.A., Ph.D., Vice Chancellor, Division of Colleges & Universities
T.K. WETHERELL, B.S., M.S., Ph.D., President of the University
LAWRENCE G. ABELE, B.S., M.S., Ph.D., Provost & Vice President for Academic Affairs
DONALD J. WEINER, B.S., J.D., Dean of the College of Law
J.B. Ruhl, B.A., J.D., LL.M., Associate Dean for Academic Affairs
DONNA R. CHRISTIE, B.S., J.D., Associate Dean for International Programs
JIM ROSSI, B.S., J.D., LL.M., Associate Dean for Research
STEPHANIE L. WILLIAMS, B.A., J.D., Associate Dean for Administration
NANCY L. BENAVIDES, B.S., J.D., Associate Dean for Student Affairs
MARK PANKEY, B.B.A., M.S., Assistant Dean for Development

FACULTY
FREDERICK M. ABBOTT, B.A., J.D., LL.M., Edward Ball Eminent Scholar in International Law
PAOLO ANNINO, B.A., M.A., J.D., Ph.D., Clinical Professor
ROBERT E. ATKINSON, B.A., J.D., Ruden, McClosky, Smith, Schuster & Russell Professor
AMITAI AVIRAM, LL.B, LL.M., J.S.D., Assistant Professor
MARGARET A. BALDWIN, B.A., J.D., Associate Professor
BARBARA ANN BANOFF, B.A., J.D., Professor
DEBORAH LYNN BASSETT, B.A., M.A., J.D., Associate Professor
TAMARA BLENKHORN, B.A., J.D., Visiting Assistant in Law (Legal Writing)
SUSAN B. BODELL, B.A., M.A., J.D., Visiting Assistant in Law (Legal Writing)
CURTIS BRIDGEWATER, B.A., PH.D., J.D., Assistant Professor
DONNA R. CHRISTIE, B.S., J.D., Elizabeth C. & Clyde W. Atkinson Professor
MARY A. CROSSLEY, B.A., J.D., Florida Bar Health Law Section Professor
TALBOT D’ALEMBERTE, B.A., J.D., President and Emeritus
VALENCIA N. DAVIS, B.A., M.A., J.D., Visiting Assistant in Law (Legal Writing)
DAVID F. DICKSON, A.B., LL.B., M.S., Ph.D., Professor Emeritus
JOSEPH M. DODGE, B.A., LL.B., LL.M., Stearns Weaver Miller Weissler Alhadeff & Sitterson Professor
CHARLES W. EHRLHARDT, B.S., J.D., Mason Ladd Professor
SALLY C. GERTZ, B.A., J.D., Clinical Professor
STEVEN G. GEY, B.A., J.D., David and Deborah Fonvielle & Donald and Janet Hinkle Professor
STEPHANIE GORE, B.S., M.S., J.D., Assistant Professor
ELWIN J. GRIFFITH, B.A., J.D., LL.M., Tallahassee Alumni Professor
KIMBERLY GRIFFA, B.S., J.D., Visiting Assistant in Law (Legal Writing)
ADAM J. HIRSCH, A.B., Ph.D., J.D., David M. Hoffman Professor
ROBERT H. KENNEDY, B.A., LL.B., Professor Emeritus
JONATHAN KLICK, B.S., M.A., PH.D., Visiting Assistant in Law
LAWRENCE S. KRIEGER, A.B., J.D., Clinical Professor
JOHN W. LARSON, A.B., J.D., Professor Emeritus
JENNIFER P. LA VIA, A.B., J.D., Visiting Assistant in Law (Legal Writing)
TAHIRI H. LEE, A.B., M.A., M.A., PH.D., Associate Professor
CHARLENE D. LUKE, B.A., J.D., Assistant Professor
DAVID L. MARKELL, B.A., J.D., Steven M. Goldstein Professor
P. GREGORY MITCHELL, B.A., Ph.D., J.D., Assistant Professor
JOSHUA M. MORSE III, J.D., Professor Emeritus
JARETT C. OELTSCH, B.A., J.D., LL.M., Professor Emeritus
KIMBERLY S. PARSONS, B.S., M.A., J.D., Visiting Assistant in Law (Legal Writing)
DAVID F. POWELL, B.A., J.D., LL.M., Associate Professor
BENJAMIN J. PRIESTER, A.B., J.D., Assistant Professor
JIM ROSSI, B.S., J.D., LL.M., Harry M. Walborsky Professor
J.B. RUHL, B.A., J.D., LL.M., Matthews & Hawkins Professor of Property
MARK B. SEIDENFELD, B.A., M.A., J.D., Patricia A. Dore Professor of Administrative Law
LOIS L. SHEPHERD, B.A., J.D., D’Alembert Professor

iv
L. ORIN SLAGLE, LL.B., Dean Emeritus and Professor Emeritus
HAROLD P. SOUTHERLAND, B.S., J.D., Associate Professor
NAT S. STERN, B.A., J.D., John W. & Ashley E. Frost Professor
RUTH E. STONE, B.S., J.D., Clinical Professor
FERNANDO R. TESÓN, J.D., LL.M., S.J.D., Tobias Simon Eminent Scholar
JOHN W. VAN DOREN, A.B., LL.B., Professor
WILLIAM VANDECREEK, B.S., J.D., LL.M., Professor Emeritus
DONALD J. WEIDNER, B.S., J.D., Professor
JOLEN RAWLS WOLF, B.A., J.D., Visiting Assistant in Law (Legal Writing)
JOHN F. YETTER, B.S., B.A., J.D., LL.M., Roberts Professor

LIBRARY STAFF

FAYE JONES, M.S., J.D., Director
ROBIN R. GAULT, B.A., M.S., Ph.D., J.D., Associate Director
ANNE D. BARDOLPH, B.A., M.S., Acquisitions Librarian
MARGARET C. CLARK, B.A., M.A., Reference Librarian
MARK S. EVANS, B.S., M.S., Computer Services Librarian
JON R. LUTZ, B.A., M.A., M.S., Electronic Services Librarian
MARY M. MCCORMICK, B.A., M.L.I.S., J.D., Head of Public Services
JANICE S. ROSS, B.A., M.S., Serials/Documents Librarian
PATRICIA K. SIMONDS, B.S., M.S., Reference Librarian
ALVA T. STONE, B.A., M.S., Catalog Librarian
FLORIDA STATE UNIVERSITY

LAW REVIEW ALUMNI ASSOCIATION

EDITOR'S COUNSEL
DuBose Ausley
Amundsen & Gilroy, P.A.
John Fagan
Theriaque & Pfeiffer

BENEFACTORS
Walter E. Forehand
Hannah, Estes & Ingram, P.A.
Lawrence A. Kellogg
J. Richard Livingston

E. John & Linda Lopez
Randall O. Reder, P.A.
Joseph Barry Schimmel

SPONSORS
Thomas Alspach
Robert C. Apgar
William E. Atwater
Timothy P. Beavers, P.A.
Lilia R. Bell
Donna E. Blanton
Marie Y. Bockwinkel
T.A. Borowski Jr.
Thomas W. Brooks
Michael P. Bruyere
Jennifer S. Buttrick
Matthew W. Buttrick
John M. Byrne
Geraldynne Carlton, P.A.
Gerald & Carol Cope
Dubow & Dubow
Joe G. Dykes Jr.
Alan B. Fields
Miranda F. Fitzgerald
Joseph R. Flood Jr.
Landis Graham French

Jody B. Gabel
Enrico G. Gonzalez
Richard E. Gordon
Gorman & Matthew, P.A.
Andrew L. Granger
Michael T. Gray
Terence A. Gross
Steven C. Hartsell
Kristie Hatcher-Bolin
Paul M. & Leslie Hawkes
Timothy P. Hoban
Thomas F. Icard Jr.
J. Isaac
Chet Kaufman
Jason Kellogg
Russel Lazega
Bruce A. Leinback
Raymer F. Maguire III
Anthony & Julie Mazzeo
Stevan D. Mitchell
John F. Pauly

Richard & Kathleen Phillips
Patrick J. Rengstl
Douglas & Laura Roberts
Robert L. Rogers III
Barbara Sanders, P.A.
Linda Carr Schmidt
Floyd & Eileen Self
Stephen R. Senn
Michael A. Sjuggerud
David R. Slaton
Karen K. Specie, L.L.C.
Marc A. Sussman
William & Connie Tarr
Molly J. Tasker
Philip & Virginia Townes
Rebecca Mercier-Vargas
Roberto Vargas
Keith & Sara White

PATRON
Kara Tollett Oakley

MEMBERS
Brett J. Berlin
Michael A. Campbell
Dennis G. Collard
Elizabeth G. Demme
Conal F. Doyle
Sally & Marc Gertz

A. Woodson Isom Jr.
Bruce Kuhse
Deborah J. Meyer
Sarah E. Spector
Susan L. Stephen
Nicole P. Wishart
General Information

Cite As. Cite this journal as FLA. ST. U. L. REV.

Submissions. The Florida State University Law Review welcomes unsolicited manuscripts on topics of interest to the legal profession. The Law Review acknowledges that unsolicited manuscripts may be submitted to other journals. All pieces will be evaluated as expeditiously as possible, but the Law Review assumes no responsibility for the return of unsolicited manuscripts. Manuscripts should be typewritten and double-spaced, with notes in footnote form. Citations should conform to The Bluebook: A Uniform System of Citation (17th ed.). Citations to Florida sources should conform to the Florida State University Law Review Florida Style Manual (6th ed.), a copy of which may be obtained for $5 from the Law Review. The Law Review accords great deference to an author’s work but reserves the right to determine the final form in which articles will be published.

Advertising. The Law Review accepts advertisements that may be of use to its readers. Rates for full- and half-page advertisements are available from the Editor-in-Chief at the mailing address and telephone number below.

Subscriptions. The subscription price is $30 for one year’s four issues or $10 for a single issue. Florida firms and individuals should add the appropriate sales tax. Foreign subscribers should add 50 cents per issue or $2 per year for postage. Subscriptions are renewed automatically unless the subscriber sends a timely notice of termination. All notifications of address changes should include old and new addresses with ZIP+4 codes. Article reprints are available for 15 cents per page by contacting the Office Manager. Back issues more than one year old may be purchased from William S. Hein & Co., 1285 Main St., Buffalo, NY 14209.

Correspondence. Address all correspondence to the Florida State University Law Review, College of Law, Florida State University, Tallahassee, FL 32306-1601. Direct manuscripts and manuscript inquiries to “Submissions.” Direct all other advertising and business matters to the Editor-in-Chief. Contact the editorial and business staff at (850) 644-2045.

Errata. Bring errors in this volume to the attention of the Editor-in-Chief. The Law Review will publish corrections when necessary.

Access. This document is available upon request in other formats for individuals with print-related disabilities. Contact the ADA Office, 301 Westcott Administration Building, Florida State University, Tallahassee, FL 32306-1470, (850) 644-5283 (Voice) or (850) 644-6693 (TDD).

The Law Review is also available on the World Wide Web at http://www.law.fsu.edu/journals/lawreview.
THE MISSION OF THE Florida State University Law Review

To provide a forum for contemporary legal discourse, we must:

Adhere to the highest standards of analysis, accuracy, and timeliness in publishing the Law Review;

Distinguish the College of Law as a leading institution, known for its support of scholarship and academic pursuit;

Attract the College of Law’s finest legal communicators by encouraging independent legal thought, offering positions of responsibility, and maintaining a tradition of service to our College;

Inform our profession about the compelling issues of our time by providing a forum for student scholars, practitioners, and legal educators;

Offer our profession a selection of articles balanced between the theoretical and the practical;

Uphold the honor and decorum of our profession by championing the highest standards of conduct;

Stand together as advocates of quality legal education and strive to improve the reputation of our Law Review and of our school;

Inspire cooperation, commitment, and camaraderie among our membership and the College of Law at large; and

Safeguard an institutional memory and tradition so that those who follow will benefit from our experience.
EMPIRICAL MEASURES
OF JUDICIAL PERFORMANCE:
AN INTRODUCTION TO THE SYMPOSIUM

STEVEN G. GEY* AND JIM ROSSI**

I. CONCEPTUAL CRITIQUES OF THE JUDICIAL TOURNAMENT: CAN WE MEASURE
JUDGES, AND IF SO, HOW? ................................................................. 1004
II. NEW EMPIRICAL WORK ON JUDICIAL PERFORMANCE: HOW DO WE
OPERATIONALIZE AND MEASURE IT? ..................................................... 1007
III. DO MEASURES OF JUDICIAL PERFORMANCE CREATE POSITIVE BEHAVIORAL
AND INSTITUTIONAL INCENTIVES? ....................................................... 1011
IV. CONCLUSION .................................................................................. 1014

In the highly charged political process of vetting, presenting, and
approving federal judicial nominees, it is commonplace for Presi-
dents, Senators, and interest groups to make claims about a nomi-
née’s merit or lack thereof. Both supporters and opponents of nomi-
nées often phrase their positions in objective terms of merit.

Independent groups such as the American Bar Association (ABA)
also appeal to seemingly objective terms of merit in evaluating judi-
cial candidates. The ABA’s ratings of judicial appointees are specifi-
cally phrased in terms of the nominee’s professional qualifications,
and in recent years the objectivity of these ratings has frequently
been the subject of controversy. In the recent dispute over the nomi-
nation of California Supreme Court Justice Janice Rogers Brown to
the U.S. Court of Appeals for the Ninth Circuit, for example, no one
on the fifteen-member ABA evaluation committee rated Brown “well
qualified,” a majority rated her merely “qualified,” while some mem-
bers of the committee rated her “not qualified.” This resulted in the
ABA committee releasing an oxymoronic rating of “qualified/not
qualified.”

Of course, the rhetoric over judicial qualifications is most
impassioned in the debate over selecting U.S. Supreme Court Jus-
tices. After two members of the ABA evaluation committee rated
Clarence Thomas “not qualified” to be appointed to the Supreme
Court, President Bush proposed ending the organization’s fifty-year
role in the judicial appointments process, thus eliminating even the

* David and Deborah Fonvielle and Donald and Janet Hinkle Professor, Florida
State University College of Law.
** Harry M. Walborsky Professor and Associate Dean for Research, Florida State
University College of Law.
1. Neil A. Lewis, Battle Lines Already Forming Against a Bush Court Selection, N.Y.
2. Neil A. Lewis & David Johnston, Bush Would Sever Law Group’s Role in Screen-
pretense that an objective assessment of judicial quality was possible.

Claims concerning judicial nominees’ professional qualifications, however, are rarely gauged against empirical evidence which might support or contradict them. In recent years, extensive data on the judiciary and individual judges has become widely available to political scientists, economists, and legal scholars, presenting a fertile opportunity for the empirical study of courts and judges. Today, the outcomes of the judicial process—judicial opinions—are widely available in searchable form through electronic databases, including Westlaw and LEXIS. More extensive common datasets for many court systems are also available to researchers.3

If you build a squash court, it does not take long for a player to find it. Indeed, given the rich availability of data, in recent years, empirical studies of judicial institutions4 and judicial behavior5 have proliferated. So have studies of judges.6 One of the more provocative recent studies of judicial behavior, by Stephen Choi and Mitu Gulati, builds on the rich availability of data on the performance of individual judges.7 Their judicial tournament, which purports to introduce calm objectivity into the study of judicial performance and judicial selection, has generated a furor of sorts. Choi and Gulati argue that

3. In 2000, Judicature, the official journal of the American Judicature Society, devoted several articles to the topic of the use of data on courts to understand the judicial process. The issue surveyed the emergence of a broad range of common datasets, covering a diverse range of courts including the U.S. Supreme Court, federal appellate courts, and many state courts. See Symposium, Social Science, the Courts, and the Law, 83 JUDICATURE 217 (2000).


the selection of Justices for the U.S. Supreme Court ought to be based on a tournament, in which judges who possess the most merit, as measured empirically, would be selected over their lower-ranked peers. If their tournament were seriously adopted, promotions to the U.S. Supreme Court would be based on quantitative measures of, rather than qualitative claims to, merit.

When Choi and Gulati first circulated their proposal, many legal scholars (including us) believed it a clever experiment of Swiftian proportions. Choi and Gulati are more than witty provocateurs. They are also hard-working, smart, and careful scholars. Since proposing the tournament, Choi and Gulati have collected and analyzed empirical data on federal appellate court judges, ranking every sitting appellate court judge over their test period. The federal judiciary is not yet setting salaries based on their tournament ranking, but their effort to operationalize and measure the performance of judges is serious business. Not only would we choose the next Supreme Court Justice from among their tournament’s high performers if Choi and Gulati’s understanding of merit were the standard, but they have done the heavy lifting for the next President to propose a nomination to the Supreme Court, actually conducting their tournament by ranking sitting appellate court judges based on measures of their productivity, influence, and independence.

Inspired by the burgeoning empirical literature on the judiciary, as well as Choi and Gulati’s proposal that empirical study (including a tournament) be used in judicial selection, the editors of the Florida State University Law Review have given Choi and Gulati a little company on the court by soliciting some essays addressing the topic of empirical measures of judicial performance from leading scholars as well as some federal judges who, willingly or not, played in their tournament. The essays in this Symposium address empirical measures of judicial performance from a variety of methodological perspectives, but they can roughly be organized around three basic themes. First, many of the essays critique the empirical enterprise itself and especially the tournament strategy for evaluating judges, although these essays also raise important issues for future empirical study of judges. Second, many of the essays in the Symposium propose new ways of operationalizing the empirical study of judicial performance or present fresh empirical evidence about judges and courts. Third, some of the essays focus on the behavioral and institutional implications of empirical studies on judges and courts.

---

I. CONCEPTUAL CRITIQUES OF THE JUDICIAL TOURNAMENT: CAN WE MEASURE JUDGES, AND IF SO, HOW?

It is certainly not new to claim that quality in judging is incapable of empirical measurement. Judge Harry Edwards of the U.S. Court of Appeals for the D.C. Circuit is perhaps the strongest critic of efforts to quantitatively measure the activities of judges and courts. For example, in a much-cited article on attempts to use quantitative techniques to study the judiciary, he observes that quantitative studies of judicial decisions “must be viewed with great caution.”9 He further states, “Regression analysis does not do well in capturing the nuances of human personalities and relationships, so empirical studies on judicial decision making that rely solely on this tool are inherently flawed.”10

Many of the essays in this Symposium echo Judge Edwards’ skepticism toward quantitative studies of the judiciary. The essays add several fresh perspectives and concerns to the mix of critics of quantitative measures of judicial performance. Some of the essays raise a concern with quantitative measures of what is fundamentally qualitative, as did Judge Edwards. For instance, one concern that reverberates throughout the various criticisms of Choi and Gulati’s tournament is that it interferes with judicial independence. Other concerns relate to the imperfections of various quantitative measures of judicial performance, suggesting that we look to different criteria than Choi and Gulati’s. In addition, some reject the extension of ranking or quantification of any sort to individual judges and judicial institutions.

Brannon Denning expresses some skepticism about an empirical judicial appointments tournament by challenging some of the assumptions underlying Choi and Gulati’s proposal.11 In particular, Denning questions the assumption that politics has overwhelmed the judicial selection process. Denning first argues that Choi and Gulati have not carefully defined the “politics” with which they are concerned. Denning argues that when Choi and Gulati refer to “politics,” what they really mean is “ideology.” Denning then critiques the claim that ideology should not enter the judicial appointments process. Denning suggests that the problem (if it is a problem) with the introduction of ideology into the process lies not with the President and individual Senators, but with the constituents they represent. If

members of the public view judging as at least in part an ideological process, Denning asks, then why should an appointee’s ideology be excluded from the range of concerns discussed during the confirmation process? After posing this question, Denning moves on to praise Choi and Gulati for providing one possible model of a more sophisticated framework for critiquing nominees during the judicial selection process. Denning concludes that something like the Choi and Gulati system could improve both the process itself and the public’s understanding of that process.

Steven Goldberg’s essay, Federal Judges and the Heisman Trophy, challenges Choi and Gulati’s basic premise that lower federal courts are the most logical places to identify candidates with the proper qualifications to be great Supreme Court Justices. Goldberg points out that most of the individuals who have become Supreme Court Justices did not serve on lower federal courts. They came, instead, from the state courts, the executive or legislative branches of government, or private practice. Even more striking, Goldberg surveys eleven lists of highly successful Justices and finds that the Justices who served on lower federal courts fared very badly on those lists. Only one of the twenty-three Justices who appear on two or more of those eleven lists had previously served on a lower federal court. Goldberg concludes by drawing an analogy between Choi and Gulati’s tournament of judges and other flawed efforts to predict greatness in other fields. In particular, Goldberg notes the failure of most Heisman Trophy winners to succeed in professional football. Goldberg suggests that success at the lower levels of the federal courts may require different talents than success at the highest judicial level, just as success in college football may require different talents than success at the professional level. Goldberg leaves open the possibility that the talents required to succeed at the highest judicial level can be empirically assessed but concludes that Choi and Gulati’s tournament is unlikely to predict greatness in Supreme Court Justices.

John Orth’s essay looks to history as a reminder that judges have been judged as long as we have had courts. Of course, elite and lay public opinion frequently is directed toward judicial decisions as well as individual judges. Many states directly elect judges, a relic of the Jacksonian Era. Litigants, too, frequently cast judgment on the judges before whom they have appeared. In addition, the political process may serve to critique judges through the process of impeachment and other removals. Despite these many institutionalized mechanisms for

evaluating judges, Orth is reluctant to embrace efforts to measure judicial performance empirically as a basis for judicial selection and promotions. To the extent that most empirical metrics focus on the outcome of courts—the decision rather than the opinion—Orth argues that empirical studies fail to capture important aspects of the judicial process and function. Moreover, Orth argues that many empirical measures of influence, such as citation studies, rely primarily on judicial peers to judge performance, thus threatening the constitutional balance of powers by promoting consensus candidates, at the cost of political judgment by other branches.

Lawrence Solum, who was one of the first in print to criticize Choi and Gulati’s tournament idea,14 poses a serious conceptual challenge to any empirical tournament of judges.15 As he suggests, by beginning with the available data rather than a concept of judicial excellence, Choi and Gulati beg a fundamental question: What is judicial excellence? Solum begins his project with virtue rather than measures of virtue. Solum identifies the “thin” judicial virtues—those on which there is widespread agreement—as extending much broader than Choi and Gulati’s measures. These virtues include incorruptibility and judicial sobriety, civic courage, judicial temperament and impartiality, diligence and carefulness, judicial intelligence and learnedness, and craft and skill. In addition, Solum makes a plea for recognition of “thick” judicial virtues: nomimos (roughly, the virtue of justice) and phronimos (roughly, judicial wisdom). If, as Solum suggests, these virtues in the practice of judging are what constitute excellence, then, clearly, it cannot be measured empirically. Solum further observes that to the extent Choi and Gulati measure something less than excellence, he sees many opportunities for gaming this already imperfect measure. Solum argues that the tournament fails to produce accurate and meaningful statistics, let alone a useful and legitimate process of judicial selection based on excellence. Nevertheless, he sees the judicial tournament as a useful thought experiment to the extent that it asks valuable questions about both the content of judicial excellence and the selection of judges who possess it.

David Vladeck brings a litigator’s perspective to bear to the critique of an empirical tournament as a measurement of judicial performance.16 Vladeck shares Choi and Gulati’s distaste for a political judicial appointments process and litmus tests as a basis for judicial selection, suggesting that litigators look for competent judges who value collegial-

---


ity and efficiency in dispute resolution. Most litigators would rather be spared the “inept, plodding, or even mediocre” judges, even if those judges are seen as the champions of important political values. Vladeck is wary of efforts to attribute objectivity and quantitative measurement to the judicial attributes that litigators value. For example, he observes that focus on the product of the judicial decision fails to capture much of the collaborative process of judging. Yet he does see some modest promise to the enterprise of empirical measurement of judges as delimiting a baseline for determining judicial competence in judicial selection. Vladeck would broaden Choi and Gulati’s factors to include many questions relating to the experience of a potential nominee in practice, government, and the community. He would see the absence of empirical performance on the relevant criteria as a danger signal, but superior performance on one or more criterion should not necessarily be a determinative factor in influencing judicial selection.

II. NEW EMPIRICAL WORK ON JUDICIAL PERFORMANCE: HOW DO WE OPERATIONALIZE AND MEASURE IT?

Choi and Gulati’s study poses many new questions as to how we should best measure variables empirically. Several of the essays in this Symposium refine the operationalization of the criteria Choi and Gulati put forth. Others attempt new measurements of judicial performance or put Choi and Gulati’s criteria to a reality test against the data. These essays add important knowledge to the body of empirical scholarship on courts and judges.

James Brudney starts his essay with an assessment of how effectively Choi and Gulati’s proposed tournament would have predicted the careers of two Supreme Court nominees from an earlier era. Brudney applies the Choi and Gulati analysis to the appointments of Chief Justice Burger and Justice Blackmun. Brudney concludes that the Choi and Gulati analysis would have produced inaccurate predictions about their Supreme Court careers. Under the Choi and Gulati analysis, Chief Justice Burger should have done very well on the Supreme Court. According to the Choi and Gulati criteria, Burger scored favorably on productivity and independence measures while serving as a court of appeals judge, while Justice Blackmun scored markedly lower in both areas. In contrast to this prediction, Burger is widely viewed as a poor Chief Justice, while Blackmun is equally widely viewed as relatively distinguished. Brudney concludes that Choi and Gulati’s criteria fail to measure intangible personal factors that contribute to greatness in Supreme Court Justices—such as the

17. Id. at 1416.
arrogance and aloofness that characterized Chief Justice Burger’s tenure on the Court and the ability to change and grow that characterized Justice Blackmun’s Supreme Court career. Brudney also draws the broader conclusion that Choi and Gulati’s system unwisely attempts to remove policy and ideological considerations from the judicial selection process. Brudney notes that many observers (and members of the general public) view these factors as highly significant, especially in light of the Supreme Court’s important role in shaping the basic collective values of the American political system.

Stephen Choi and Mitu Gulati continue their effort to devise objective measures of quality and judicial merit in their contribution, Which Judges Write Their Opinions (And Should We Care)? They use techniques from computational linguistics and other methods to explore both the desirability and feasibility of determining whether individual judges write their own judicial opinions. In the first part of their essay, Choi and Gulati argue that knowing the authorship of judicial opinions is highly relevant when deciding whether to elevate a judge to a higher court. They also argue that this information will be useful in assessing the ongoing performance of sitting judges and in making determinations about the allocation of judicial resources. Assuming this information is useful, Choi and Gulati attempt to devise a method to assess judicial authorship. Unfortunately, the methodology chosen by Choi and Gulati failed to identify judges who, by reputation, are known to write their own opinions. Choi and Gulati speculate on the failure of this methodology and suggest possibilities for future research using a more finely honed version of the methodology, which would control for opinions regarding different subject matter and would focus on factors such as citation practices; opinion, paragraph, and sentence length; and other, frequency-related measures.

Sharing Brudney’s views, Lee Epstein, Jeffrey Segal, Nancy Staudt, and René Lindstädt are skeptical about the broad assumptions about the judicial nomination process that motivate Choi and Gulati’s proposed tournament of judges. Specifically, these four authors dispute the conclusion that policy determinations and ideological concerns have come to dominate the process of judicial nomination and confirmation. They base this claim on their comprehensive study of all votes cast by individual Senators on Supreme Court nominees since Earl Warren in 1953. Their study indicates that although ideology plays a significant role in determining whether a


Senator will vote for or against a nominee, the qualifications of the nominee will play virtually the same role. It is the interplay of these two factors—ideology and qualifications—that provides the best predictor of a Senator’s vote. While Senators will almost certainly vote for a nominee whose ideology is similar and who is highly qualified and will almost certainly vote against a nominee who is ideologically dissimilar and unqualified, there are certain conditions in which Senators will often vote for an ideologically dissimilar nominee who is highly qualified. The authors conclude that despite common complaints about the nomination system, the qualifications of a nominee continue to play a significant role in the confirmation process. The authors argue that this conclusion should give pause to those—including Choi and Gulati—who propose to fundamentally alter the current judicial selection process.

Daniel Farber starts his essay, *Supreme Court Selection and Measures of Past Judicial Performance*, by praising Choi and Gulati’s contribution to the judicial selection literature. Farber specifically compliments them for showing how objective measures of past judicial performance could improve the judicial selection process. He then suggests, however, that Choi and Gulati’s proposals are far too ambitious based on the evidence they have mustered. Farber argues that we should be reluctant to adopt their proposal because it is impossible to measure past judicial performance as precisely as they claim and because professional merit should not be the only factor in the judicial selection process. With regard to the first point, Farber notes that some of the criteria used by Choi and Gulati to measure productivity and independence are imperfect because high scores on these criteria may be a function of a judge’s personal characteristics (such as self-centeredness) that would not well serve a Supreme Court Justice. Other measures, such as the use of citation counts as a proxy for influence, are imperfect because of outlier effects and feedback loops that generate citations primarily due to a judge’s personal prominence, which may or may not be directly correlated with objective merit. In the end, Farber engages Choi and Gulati on their own terms more than most of the commentators—to the extent their tournament provides at least a rough measure of judicial merit—but he also argues that merit alone should not be (and probably will never be) the deciding factor in judicial selection.

Michael Gerhardt’s essay surveys the problems with defining three critical factors that are often central to the judicial selection

---
process: merit, the “mainstream,” and ideology. In the first part of his essay, Gerhardt documents the different ways in which merit is defined by different constituencies and discusses the assumptions about the judicial role that underlie these different definitions. The second part of Gerhardt’s essay investigates the effort to define the “mainstream” in battles over judicial appointments. Although the fight over the “mainstream” is common to almost all judicial appointments battles, it is unclear how this middle ground should be defined. Gerhardt provides several options and considers the various perspectives from which the middle ground should be viewed. The third part of his essay assesses the ways in which the effects of ideology can be assessed in judicial appointments battles. Despite the difficulties presented by these three aspects of the judicial appointments process, Gerhardt argues that an objective empirical study of merit is both possible and desirable. In the absence of some mechanism to assess judicial appointments by reference to a coherent and comprehensive definition of merit, Gerhardt concludes, judges will be viewed as little more than “policymakers who just happen to wear robes.”

Michael Solimine’s contribution, Judicial Stratification and the Reputations of the United States Courts of Appeals, advances the empirical project by applying the tournament concept to performance across, rather than within, appellate courts. Like many of the other contributions, Solimine disentangles the concepts of reputation, prestige, and influence in hopes of identifying a dependable measure of quality. Insofar as the subject of empirical inquiry is federal appellate court judges, he suggests that an effort to separate empirical study of the performance of judges from the performance of the courts they compose is useful. After surveying historical efforts to measure the reputations of courts of appeals, in which the D.C. Circuit and Second Circuit seem to consistently enjoy the highest reputation and most influence, Solimine considers how studies of the citation analysis of particular appellate court judges relate to the reputation of the circuit on which they sit. Some of the measures of influence, Solimine observes, place the Seventh Circuit above the D.C. Circuit and Second Circuit. Solimine’s analysis of the rise and fall of reputation sheds light on the nature and importance of reputation of appellate courts generally. For instance, he observes that during the latter part of the twentieth century, reputation among circuits became homogenized and the importance of reputation fell among litigants.

23. Id. at 1235.
Solimine’s analysis also illuminates disconnects in reputation between appellate circuits.

III. DO MEASURES OF JUDICIAL PERFORMANCE CREATE POSITIVE BEHAVIORAL AND INSTITUTIONAL INCENTIVES?

Regardless of whether and how we measure judicial performance, everyone agrees that the empirical enterprise has important implications for behavioral incentives and institutions. If we measure—and especially if we rank or otherwise evaluate—actors in an institutional setting based on their performance on a set of criteria, these actors will adjust their conduct in response to the positive and negative incentives created by this information. This seems uncontroversial in concept. But any evaluation of judicial incentives requires some specification of the motivations of judges. In a famous article, Judge Richard Posner suggested that judges maximize “the same thing everybody else does,” but among (most) lawyers and judges, it remains controversial to claim that judges are motivated by anything other than doing justice. For political scientists, economists, and some legal scholars, discussion of incentives affecting judicial behavior is much less maligned. Several of the essays make important behavioral and institutional insights which will have implications for further empirical and conceptual research on judges and courts.

Judge Jay Bybee and Thomas Miles join other commentators in questioning Choi and Gulati’s central argument that an empirical tournament can settle disputes over the qualifications of candidates for judicial appointments. While acknowledging that empirical assessments are an important factor in the appointments process, Bybee and Miles argue that other, less quantifiable subjective measures of quality are even more significant. For example, they suggest that an analysis of a candidate’s position on “hot-button” issues such as affirmative action and abortion can often reveal more about a candidate’s qualifications than the empirical measures suggested by Choi and Gulati. Also, Bybee and Miles observe that the tournament would create strong incentives for ambitious judges to engage in undesirable behavior simply to increase their judicial score. Ambitious judges will be encouraged to “judge to the tournament,” which will have the effect of undermining the spirit of judicial independence and pursuit of judicial quality that the tournament is supposed to encourage.

27. Id. at 1068-73.
Judge Richard Posner’s contribution to the issue, *Judicial Behavior and Performance: An Economic Approach*, should be taken very seriously by both judges and scholars—and not only because Posner is the likely victor in Choi and Gulati’s tournament of appellate judges. As is characteristic of Posner’s work, the contribution is brimming with valuable insights about the implications of empirical measures of individual performance of judges. Posner challenges empirical and behavioral scholars of the judiciary to start with a more serious institutional framework for understanding decisions rather than simply with data. As Posner’s approach would suggest, “judicial behavior is likely to differ across national legal systems and indeed within a nation’s legal systems to the extent that components of the system . . . differ in the incentives and constraints that they impose on judges.” He extends the analysis not only to appellate court judges—for which he, like Brudney, would like to see greater historical research of how a judge’s performance on the U.S. court of appeals maps onto performance as a Justice of the U.S. Supreme Court—but also to the different institutional settings of U.S. district court judges, state court judges, and arbitrators. Posner illuminates how each of the institutional settings presents unique behavioral incentives for individual judges. His analysis of behavioral incentives in the institutional context raises many additional questions and hypotheses for empirical research not only of the behavior of judges but also of other actors within the judicial system.

Judge Bruce Selya evaluates the dual objective of the tournament: merit-based evaluation and increased incentive to perform. The proliferation of multiple objective metrics may undermine the goal of political transparency: he suggests that “any ranking system will face constant criticism that it is a proxy for either political affiliation or ideological leanings rather than for merit.” As to Choi and Gulati’s criteria, Judge Selya takes issue with several of them—particularly their incentive effects. Commenting on the manipulability of citation measures, he suggests that “[a]ny judge worth his salt will tell you that there are ways to write opinions that make citation more likely.” He also characterizes the measurement of judicial independence by use of dissents as useless as well as perverse in its incentives. Selya points out that, over time, “judicial rankings will say less about actual merit and more about agility—the ability to game

29. Id. at 1259.
31. Id. at 1286.
32. Id. at 1290.
the system.” The inevitable result of these incentives, Judge Selya warns, is to undermine the very goals of the tournament.

Russell Smyth’s essay illustrates the promise of taking seriously the institutional context of judging. Smyth rejects the suggestion of Judge Harry Edwards that the collegiality of judging cannot be measured, suggesting that “there is much evidence of the success of regression analysis in capturing these nuances of human behavior.” Smyth examines how the idea of the judicial tournament would translate in the setting of judicial selection in Australia. He carefully builds from an evaluation of judicial incentives, pointing out both behavioral and quantitative problems with many of the Choi and Gulati criteria along the way. Smyth argues that a tournament transplanted to the Antipodes would best focus on minimum qualifications for judges, rather than on the highfliers. His essay illustrates how one need not reject empirical study of judges in order to accept the criticisms of the judicial tournament many others make.

Ahmed Taha’s essay provides a limited defense of an empirical judicial tournament. Taha focuses on the effects of rankings on the judicial nomination and confirmation process as well as the effects of rankings on judges’ behaviors. He argues that the case for a tournament of judges as a judicial selection device is stronger for the positions to which Choi and Gulati apply the tournament—U.S. courts of appeals—than for positions on the U.S. Supreme Court. As he argues, basing selection of U.S. Supreme Court Justices on a tournament may undermine Choi and Gulati’s goals by allowing “a President to nominate more politically extreme candidates who happen to have high merit rankings.” By contrast, “a ranking of federal district [court] judges would avoid many of the problems that might be created by ranking appellate judges.” Further, based on an assessment of judicial behavior, he suggests that, if subjected to the tournament, federal district judges are more likely to respond to the positive incentives it creates.

33. Id. at 1294-95.
35. Edwards, supra note 9, at 1640-43, 1656.
37. See supra Part I.
39. Id. at 1406.
40. Id. at 1403.
IV. Conclusion

There is an elegance about the empirical tournament as a mechanism for making the intangible knowable and defusing political rhetoric in judicial selection. Of course, it does not succeed in these respects. As the essays in this Symposium suggest, the project of empirical evaluation of judicial performance must also address normative questions of quality, and it raises important issues about measurement, behaviors, and incentives. A focus on empirical issues relating to the activity of judging and its outputs not only feeds the prurience of the legal community for gossip about what lies beneath the robes of judges, but it can help to shed light on fundamental issues of judicial behavior, important to students and scholars in law, political science, and economics. It also challenges scholars to disentangle important questions relating to the quality of judging and judicial institutions. For example, to what extent can judicial performance be reduced to a judge’s reputation? Or does quality in judging represent something more? To what extent, if any, does judicial quality in one institutional setting, such as in the context of the U.S. circuit courts of appeals, necessarily translate into quality in another judicial setting, such as in the context of the U.S. Supreme Court? Does a Supreme Court nomination system that keeps Judge Posner from becoming Justice Posner really fail to recognize or reward merit?

Choi and Gulati’s empirical tournament presents a much-welcomed challenge: How might we introduce greater objectivity into discussions of merit in judicial selection? The scholarly ruminations in the pages of this Symposium certainly cannot answer all of the questions empirical studies of judging and courts present. Much as the initial furor the tournament provoked, some of the commentators in this Symposium refuse Choi and Gulati’s challenge on normative terms. Nevertheless, as the Symposium contributions clearly indicate, an empirical tournament tells us much about several enterprises: normative theories of judging, quality, judicial independence, and judicial selection; empirical discussions of measurement of judges and courts; and behavioral and incentive-based evaluations of institutions. As the essays in the Symposium indicate, the empirical tournament has inspired some important advances in the discourse about measurement of performance in the context of the judiciary and its relevance to the selection of judges and the judicial process. That is a discourse that will be certain to continue as long as we have data, judges, and courts.
In two recent, provocative articles, Professors Stephen Choi and Mitu Gulati contend that a tournament based on objective considerations of judicial merit should govern our approach to the nomination and confirmation of Supreme Court Justices.1 Professors Choi and Gulati developed their three quantitative tournament criteria—seeking to measure productivity, independence, and quality among sitting appellate judges2—for prospective application. It seems reasonable that these same criteria could be used to compare two contemporaneous Supreme Court nominees from a somewhat earlier era, in order to consider whether one emerges as more “worthy.” Such a comparison is likely to be especially instructive if both nominees then ended up serving on the Supreme Court, thus allowing for

2. Choi & Gulati, Tournament, supra note 1, at 305-10; Choi & Gulati, Empirical Ranking, supra note 1, at 42-67.
observations regarding their actual performance, not simply their tournament-related potential.

Two such Justices conveniently exist: Warren Burger and Harry Blackmun. As federal appellate judges, Burger and Blackmun served for comparable lengths of time during the same historical period. They were nominated for the Supreme Court by the same President, who had made clear that he wanted new members of the Court to reflect a certain judicial philosophy. As Supreme Court Justices, however, Burger and Blackmun came to differ sharply in their doctrinal and ideological orientation. Commentators also have diverged in evaluating their respective careers on the Court.

In this Essay, I compare the “objective merit” outputs of Warren Burger and Harry Blackmun as appellate judges, in order to consider some of the challenges involved when assessing the merit-related potential of Supreme Court nominees. Part I discusses the appellate court records of Judges Burger and Blackmun, borrowing from Choi and Gulati’s three quantitative criteria. Judge Burger’s performance appears more promising in the area of productivity and on one measure of independence, while the two judges seem comparably strong with respect to the quality factor. Part I then suggests how little guidance these quantitative assessments provide when reviewing the two men’s careers on the Supreme Court. For Justice Burger, a record of independence in appellate court opinion writing seems, in retrospect, to have foreshadowed an aloof and at times fractious attitude toward his colleagues while serving as Chief Justice. Justice Blackmun’s tenure on the Court was characterized by his evolving perspectives—in terms of the social vision he embraced and the effect that vision had on his judicial philosophy. Blackmun’s evolution, which is often invoked by persons who consider him a distinguished Justice, could scarcely have been anticipated based on a quantitative review of his appellate court performance.

Part II addresses in more general terms certain reservations about the performance measurement approach proposed by Choi and Gulati. Efforts to assess potential judicial merit are surely necessary prerequisites to appointment, but they ought not to preclude consideration of political and ideological factors. The Constitution contemplates that a candidate’s partisan or ideological background may be


part of the selection process, and the political branches’ reliance on this background reflects an understanding that the Supreme Court’s judgments embody important choices on matters of public policy as well as on the rule of law. In addition, when attempting to evaluate judicial merit, we should recognize the relevance and importance of nonquantitative factors, including collegiality and career diversity. The Choi and Gulati approach inevitably overlooks or undervalues such qualitative elements.

I. COMPARING POTENTIAL AND REFLECTING ON PERFORMANCE

Professors Choi and Gulati have identified a set of objective measures they would like to see employed when comparing appellate judges for potential elevation to the Supreme Court. These measures focus on three categories: productivity in generating impressive numbers of published opinions, quality of written opinions as reflected in frequency of citations outside one’s own circuit, and independence from the views of one’s colleagues and political sponsors as manifested through patterns of dissents and concurrences authored.5 While recognizing that the assignment of proper weights within and between these categories presents some challenges, Choi and Gulati maintain that the three measures should at minimum form a valuable basis for evaluating claims of merit made by Presidents on behalf of their Supreme Court nominees.6

Rather than critique the details of the Choi and Gulati approach, I borrow from their three categories to develop a comparison that better suits my focus on two judges during an earlier time period.7 I make some minor adjustments or refinements in the three measurement categories,8 while attempting to be faithful to the spirit of the Choi and Gulati enterprise.

5. Choi & Gulati, Empirical Ranking, supra note 1, at 42-43, 48-50, 61-63 (discussing the three measures).
6. See Choi & Gulati, Tournament, supra note 1, at 310-11; Choi & Gulati, Empirical Ranking, supra note 1, at 29-30.
7. In general, I have examined the records of Judges Burger and Blackmun over a greater number of years than the three year time frame relied on by Choi and Gulati, and have focused only on the circuits in which Burger and Blackmun served during that longer period. Given time and space constraints, I cover far fewer judges than do Choi and Gulati, and my treatment of their three factors is less elaborate.
A. The Output-Based Potential of Judges Burger and Blackmun

Warren Burger and Harry Blackmun had remarkably similar backgrounds before ascending to the Supreme Court. Each was raised in modest financial circumstances in St. Paul, Minnesota, where they were childhood friends. Each came from a moderately conservative Protestant, Republican family tradition. Each was chosen by President Eisenhower to join the federal appellate bench. After more than a decade of appellate court tenure, each was nominated by President Nixon and then confirmed to serve on the Supreme Court.

When their performances as appellate court judges are considered under the Choi and Gulati criteria, Burger appears to have somewhat more potential than Blackmun. The fact that Blackmun is generally regarded as a more distinguished Supreme Court Justice may not have been foreseeable at all. Yet, to the extent that hindsight offers some perspective regarding the reasons for their divergence, the three quantitative measures proposed by Choi and Gulati are of little help; in one instance they may well be counterproductive. If anything, it would seem that available qualitative evidence might have served as a more useful signpost of what was to come.

1. Productivity in Opinion Writing

The first Choi and Gulati category is productivity based on published opinions. I compared the number of published majority opinions authored by Judges Burger and Blackmun to the number of


13. Choi and Gulati report on both total number of published majority opinions and total number of published opinions overall, which includes concurrences and dissents. See Choi & Gulati, Empirical Ranking, supra note 1, at 44 tbl.2. I address only majority opinions here; discussion of dissents and concurrences is confined to the “independence” factor, in part to avoid overvaluing these separate opinions as integral parts of two categories. Published majority opinions were identified through a LEXIS search for the circuit court
majority opinions published by every other active status judge serving on their respective circuits for some or all of the same time periods. Thus, Burger served for twelve full calendar years (1957-1968) as an active status judge on the D.C. Circuit; Table 1 compares his productivity to that of all other active status judges who served on the D.C. Circuit between 1957 and 1968. Similarly, Table 1 compares Blackmun’s productivity as an active status judge during his ten full calendar years (1960-1969) to the productivity of all other active status judges who served on the Eighth Circuit between 1960 and 1969. Comparisons within each circuit are based on a standardized productivity score calculated for each individual judge.

on which each judge served during each calendar year in which the judge was on active status for all twelve months. Cross-checks with Westlaw revealed that LEXIS was more accurate in identifying opinion authors during this period. In developing lists of citation-receiving opinions for use in Table 2, see infra app., I cross-checked many judges’ majority opinion numbers with Westlaw and, for Judge Blackmun, with the appendix to his Supreme Court confirmation hearing record, see infra note 71.

Including published concurrences and dissents as part of a productivity measure raises an additional problem in today’s circumstances; publication of separate opinions may well be determined derivatively based on whether the majority opinion is deemed worthy of publication. See Merritt & Brudney, supra note 8, at 107 (noting that appellate courts determined not to publish almost fifteen percent of decisions that included a dissenting opinion and nearly six percent of decisions that contained a concurring opinion, for a dataset of cases decided between 1986 and 1993). Choi and Gulati may not have accounted for this problem. See Choi & Gulati, Empirical Ranking, supra note 1, at 43 (assuming that judges who author unpublished majority opinions “affirmatively do not want [them] to be used by others as precedents”).

14. See infra app., tbl.1. I confined the comparisons to judges who served for full calendar years in an effort to standardize units of measurement. For instance, Judge Burger served for parts of two other years (1956 and 1969) and published opinions during those years, but it seemed appropriate to eliminate the impact of production during unequal numbers of months and also to minimize effects associated with the initiation period or with the consequences of an unplanned promotion-related departure. My comparisons were with active status judges to avoid reliance on senior status or visiting judges sitting by designation, whose opinion assignments may have been less uniform. See James J. Brudney & Corey Ditslear, Designated Diffidence: District Court Judges on the Courts of Appeals, 35 LAW & SOC’Y REV. 565, 581-84 (2001) (describing the modest opinion-writing role of designated district judges over seven year period).

15. Standardized productivity scores were computed by taking each judge’s number of majority opinions and subtracting the circuit’s mean for that year; the result was then divided by the standard deviation. See DAVID S. MOORE, STATISTICS: CONCEPTS AND CONTROVERSIES 251-53 (5th ed. 2001) (discussing standard scores). A score of zero would indicate that the judge’s opinion-writing frequency was at the mean level. Table 1, infra app., presents the average productivity score for each judge based on the number of full-time active years he served during our period of measurement. See supra note 14. (Figures for individual years are on file with the author. Cory Smidt provided valuable assistance in calculating the standardized productivity scores.) Professors Choi and Gulati use a somewhat different approach to capturing variations within a circuit; they adjust for intercircuit differences based on the mean number of opinions published for judges of each circuit. See Choi & Gulati, Empirical Ranking, supra note 1, at 45-46. The Choi and Gulati approach was less practicable here, given my focus on only two circuits for a period well in excess of three years. See supra note 7.
Although Judges Burger and Blackmun published majority opinions at similar rates, Burger’s relative standing on productivity among his D.C. Circuit colleagues is considerably higher than Blackmun’s in the Eighth Circuit. Judge Burger authored 254 published majorities from 1957 to 1968, or 21.17 per year. His standardized productivity score of 0.63 ranked him second overall among the fourteen judges serving during this period and first in comparison with judges who were on active status for the same number of years.\footnote{See infra app., tbl.1. Judge Leventhal, the only judge with a higher score, served only three years during this twelve year period. Notably, in six of the twelve years, Burger produced majority opinions at a number more than one standard deviation above the circuit mean.}

Judge Blackmun produced 210 majority opinions from 1960 to 1969, or 21 per year. However, Blackmun’s performance was merely average in relation to his circuit court peers. Blackmun’s standardized productivity score of 0.01 ranked him sixth out of the twelve judges serving during this period and last among the judges who served for the same number of years that he did.\footnote{See infra app., tbl.1. Unlike Judge Burger, Blackmun’s productivity was never more than one standard deviation above the circuit mean during his ten year period.}

Appellate courts published all majority opinions authored during the period in which Judges Burger and Blackmun served, and the differences in productivity among judges within each circuit appear to be less dramatic than those identified by Choi and Gulati.\footnote{See Choi & Gulati, Empirical Ranking, supra note 1, at 86-89 tbl.B (noting that in Seventh Circuit from 1998 to 2000, Judge Posner published 254 majority opinions and Judge Manion published 102). Using the ten Seventh Circuit judges in Choi and Gulati’s tournament as a baseline, Posner’s output exceeded the circuit mean of 168 published majorities by 51% while Judge Manion’s total of 102 published majorities was 39% below the circuit mean. See id. For our comparable 1964-1966 period, the top judge in the D.C. Circuit exceeded his circuit mean by 23% while the bottom judge was 25% below the mean; in the Eighth Circuit, the top judge exceeded that circuit’s mean by 16% while the bottom judge fell 7% below the mean. (Copies of all calculations are on file with author.)} This narrower range presumably reflects in part the tendency to distribute opinion assignments on a relatively equitable basis among three-member panels. When four out of every five circuit court panel decisions are \textit{unpublished}—as is true today—there is likely to be far more variation in individual judges’ rates of publication within their own circuit. Of course, a major reason for this variation may well be the opinion assignment practices of a panel’s most senior or “ranking” member. By retaining the few publication-worthy cases for themselves, these panel members can influence productivity figures in ways that Choi and Gulati may need to anticipate.\footnote{Given the traditional practice that the senior active status member of each panel controls opinion assignments, it seems worth exploring whether judges who most often have the assignment power tend to produce unusually high numbers of published opinions within a circuit. See J. Robert Brown, Jr. & Allison Herron Lee, Neutral Assignment of
Choi and Gulati’s focus on only three years of opinion writing probably exaggerates productivity differences among judges when compared with the ten and twelve year periods examined here.20 Nonetheless, even in an earlier era of universal publication, and measuring for productivity over a longer period of time, some judges were distinctly more productive than others within their own circuit. In this regard, Judge Burger was unusually prolific in generating majority opinions when compared to others on the D.C. Circuit. By contrast, Judge Blackmun was not unproductive, but his performance on the Eighth Circuit qualifies him as no better than average.

2. Quality of Opinion Writing

Choi and Gulati recommend citation counts as a proxy for the quality of the appellate judicial product.21 They contend that judges whose opinions help explain the law more clearly or effectively, or who develop a reputation for quality analysis outside their own circuit, will receive more citations.22 Borrowing from their approach, I compiled the citations received for majority opinions published by Judges Burger and Blackmun during a four year period when they were both serving on the appellate bench, from 1963 to 1966. I also compiled citations for the four other D.C. Circuit judges who were on active status during all four of those years and for the three other Eighth Circuit judges who served throughout the same four year pe-

---

20. Productivity is more likely to be volatile within a three year period; average production tends to flatten out over a longer time frame. For example, Judge Burger’s productivity varied over his D.C. Circuit career—it was as high as 26.7 majorities per year from 1960-1962 and as low as 14 per year from 1964-1966. It is not unreasonable to infer that Judge Leventhal’s three year productivity average of 29.0, see infra app., tbl.1, might exceed his performance when measured over a more extended period.

21. Choi & Gulati, Tournament, supra note 1, at 396-97; Choi & Gulati, Empirical Ranking, supra note 1, at 49-50.

22. Choi & Gulati, Empirical Ranking, supra note 1, at 48-50. Opinion assignment practices may affect citation counts just as they do publication rates. See supra note 19 and accompanying text. Judges who tend to control the assignment of majorities can influence reputations by channeling certain hot doctrinal subjects, or cutting-edge legal questions, to particular colleagues—or by retaining those opinions for themselves. Choi and Gulati have not controlled for “pretournament inherent differences” that they view as related to merit. Choi & Gulati, Empirical Ranking, supra note 1, at 49 (discussing possible citation advantage for judges who are better liked, on more respected circuits, or have more seniority). It would be useful to consider whether the power to control opinion assignments should be deemed merit-related in this context.
riod. For each of these judges, I gathered the total number of citations by any state or federal court outside of their own federal circuit, including citations by the Supreme Court. In addition, I compared the top twenty citation-receiving opinions of each judge; this tracks the Choi and Gulati approach of evaluating a judge’s quality based on her “best opinions” as well as her average performance when considering the full volume of production.

As presented in Table 2, Judges Burger and Blackmun are remarkably similar in terms of their volume of outside citations. For the 1963-1966 period, Judge Burger’s average number of citations per opinion based on all published majority opinions modestly exceeded Blackmun’s, 6.09 to 5.66. Conversely, with regard to citations per majority opinion for the twenty most frequently cited opinions, Blackmun’s average was slightly above Burger’s, 16.5 to 16.0.

Further, both Burger and Blackmun stand out from other members of their respective circuits in terms of the recognition garnered for their majority opinions. Judge Blackmun’s top twenty average of 16.5 far outpaces his Eighth Circuit colleagues; it is more than twice that of Judges Matthes and Van Oosterhout and nearly double that of Judge Vogel. Judge Burger is comparably impressive in the company of his D.C. Circuit colleagues. Burger’s top twenty average of 16.0 is nearly three times that of Judge Danaher, and his top twenty and overall averages are well above the citation counts for his three other colleagues. Moreover, in distancing himself from Judges Bazelon and Wright, Burger’s performance on citation counts ranks him comfortably ahead of two nationally renowned appellate judges of the era.

23. Although time constraints made it impracticable to compile lists of outside-circuit citations for all individual judges, the judicial colleagues selected provide a suitable intracircuit framework for discussion.

24. Sara Sampson, research librarian at The Ohio State University Moritz College of Law, compiled the citation counts, using “custom restrictions” within LEXIS’s Shepard’s Service to provide individual cases. The “outside circuit” citations include cases cited by the Supreme Court but do not include federal cases within a judge’s own circuit—that is, no district court cases from the D.C. Circuit are included for Judge Burger or his four colleagues and none from districts within the Eighth Circuit for Judge Blackmun or his three colleagues. Citations were to all outside opinions through May 31, 1969. This parallels the Choi and Gulati approach of tracking 1998-2000 opinions cited through May 31, 2003, Choi & Gulati, Empirical Ranking, supra note 1, at 50, although I have included four years of majority opinions instead of only three. The date of May 31, 1969, is also convenient because Judge Burger was nominated to the Supreme Court in late May; any cites after May 31, 1969, may reflect in part the perception of Burger, and eventually Blackmun, as Supreme Court Justices.

25. Choi & Gulati, Empirical Ranking, supra note 1, at 54.

26. See infra app., tbl.2.

27. See infra app., tbl.2. Blackmun’s overall citation average also far exceeds the averages for his three colleagues who served as active judges throughout this same four year period.

28. See infra app., tbl.2.
An additional issue is whether any difference existed between the two circuits in terms of their outside recognition. Based on this admittedly selective sample, it appears that D.C. Circuit judges are cited more often by courts outside their own circuit.\textsuperscript{29} There are a number of reasons why the D.C. Circuit might have enjoyed such an advantage during this time period. From a subject matter standpoint, that court has long been the primary venue for judicial review of agency action, handling an unusually large volume of administrative law cases.\textsuperscript{30} During the 1960s, it also had jurisdiction to review general criminal matters under the D.C. Code.\textsuperscript{31} Perhaps judges from other circuits were more likely to look to the D.C. Circuit as a source of doctrinal insight in the administrative law and criminal law areas, something less obviously available from the Eighth Circuit.\textsuperscript{32} It also is possible that judges elsewhere in the country simply perceived the D.C. Circuit as “higher status” in general terms and looked more often to its majority opinions for guidance.\textsuperscript{33} Still, under the Choi and Gulati model, service on a more respected circuit counts as a positive factor; it is the circuit’s reputation for high quality work, and the individual judge’s ability supporting such a reputation, that warrant preferential consideration in the tournament.\textsuperscript{34} Judge Burger’s top

\textsuperscript{29} Omitting Judge Blackmun, the Eighth Circuit average for citation-receiving opinions among judges on active status for these four years is 3.08 citations per opinion considering all majorities and 8.53 per opinion considering top twenty majorities. Excepting Judge Burger, the D.C. Circuit average for active status judges is 3.49 citations per opinion considering all majorities and 9.21 per opinion considering top twenty majorities. Also revealing is the fact that three of the four D.C. Circuit judges (besides Burger) approached or exceeded ten citations per opinion for their top twenty opinions, whereas only one of three Eighth Circuit judges (other than Blackmun) even approached that level. See infra app., tbl.2.


\textsuperscript{32} Of Judge Burger’s top twenty citation-receiving opinions in the 1963-1966 period, seven were administrative law decisions and thirteen were criminal law opinions.


\textsuperscript{34} See Choi & Gulati, Empirical Ranking, supra note 1, at 49. Choi and Gulati also consider the possibility that high citation rates may reflect the controversial or outrageous nature of an opinion rather than the quality of its analysis. Id. at 54-55. They recommend examining negative citations to the top twenty citation-receiving opinions as a safeguard. Id. at 55-57. There are virtually no such negative citations by courts for the Burger and Blackmun majorities published during our four year period. Using LEXIS’s Shepard’s Ser-
performance on such a prestigious court may therefore be worthy of extra respect.

Finally, one intriguing aspect of Judge Blackmun’s record on citations is that five of his top twenty citation-receiving opinions involve the tax field. Blackmun practiced tax law for many years before joining the federal bench, and the widespread recognition for his tax law decisions presumably reflects the value of that pre-judicial experience. Blackmun’s Supreme Court career also distinguished him as a highly respected voice on federal tax law issues; in this content-specific respect, his appellate court performance may have signaled Supreme Court potential.

3. Independence in Opinion Writing

Choi and Gulati maintain that a judge’s willingness to disagree publicly with her colleagues on the appellate bench, especially colleagues who are presumptively like-minded, is a valuable predictor of judicial independence. They identify two distinct components of judicial independence that deserve assessment. As a measure of special intellectual effort expended and simple willingness to differ from one’s colleagues, they value the number of separate opinions (dissents and concurrences) that a judge authors. As an indicator of ideological autonomy, they place additional emphasis on these dissents and concurrences if the judge’s “separation” was from panel members of her own political party.

Table 3 indicates that Judge Burger wrote 130 dissents and concurrences, far more than the eighteen authored by Judge Blackmun vice—searching for questioned, criticized, overruled (wholly or in part), or disapproved—resulted in identification of three negative cites for Burger majorities, only one of which was outside the circuit, and three negative cites for Blackmun majorities, two from outside the circuit.

35. See Gen. Bancshares Corp. v. Comm’r, 326 F.2d 712 (8th Cir. 1964); Hamm v. Comm’r, 325 F.2d 934 (8th Cir. 1963); Estate of Peyton v. Comm’r, 323 F.2d 438 (8th Cir. 1963); Banks v. Comm’r, 322 F.2d 530 (8th Cir. 1963); Bouschor v. United States, 316 F.2d 451 (8th Cir. 1963).


37. Id. at 110 (discussing Blackmun’s extraordinary reputation in this area among tax lawyers and academics).

38. Choi & Gulati, Tournament, supra note 1, at 310; Choi & Gulati, Empirical Ranking, supra note 1, at 62.


40. Id. at 63. Choi and Gulati also consider dissents written against the judge in order to assess in a more refined way the extent to which each judge opposed a judge of the same political party. See id. Their evaluations of independence based on dissents and concurrences are more extensive than what I am able to present here, although I do discuss dissents from Burger and Blackmun majorities. See infra note 48 and accompanying text.

41. See infra app., tbl.3.
during comparable periods on the appellate bench. The large differential must be understood in the context of their distinct circuit cultures. During this period, the D.C. Circuit was a contentious place; judges wrote dissenting or concurring opinions much more frequently than was true for their colleagues on the Eighth Circuit. It may therefore be appropriate in one sense to view Burger and Blackmun as displaying comparable independence within their respective circuit court spheres. Each judge authored more dissents and concurrences than the average for his circuit as a whole: Burger ranked fourth among fourteen D.C. Circuit judges in separate opinions written per year of service, while Blackmun placed fifth among twelve Eighth Circuit judges on this scale. Still, the six-fold difference between Burger and Blackmun in annual number of separate opinions reflects a considerably greater investment of intellectual effort on Burger’s part. Circuit norms may help account for the size of this gap, but the differential itself is tangible and robust.

As Choi and Gulati recognize, circuits may vary considerably among each other or over time in terms of their partisan composition, making it difficult to assess an individual judge’s autonomy from cir-

42. See infra app., tbl.3 (indicating that D.C. Circuit judges on average wrote dissents nearly seven times as often as Eighth Circuit judges (6.32 per “judge year” versus 0.93) and that they authored concurrences nearly eight times as often (3.20 per “judge year” versus 0.42)). The contentiousness among D.C. Circuit judges becomes even clearer when considering the fact that the Eighth Circuit judges actually produced slightly more majority opinions on average than their D.C. Circuit counterparts. See infra app., tbl.3 (indicating that Eighth Circuit judges produced 20.5 majorities per “judge year” compared to 18.1 majorities per “judge year” for D.C. Circuit judges). For a more detailed treatment of the unusually high levels of conflict on the D.C. Circuit during this period, see Charles M. Lamb, A Microlevel Analysis of Appeals Court Conflict: Warren Burger and His Colleagues on the D.C. Circuit, in JUDICIAL CONFLICT AND CONSENSUS: BEHAVIORAL STUDIES OF AMERICAN APPELLATE COURTS 179 (Sheldon Goldman & Charles M. Lamb eds., 1986) (analyzing effects of party and religion in explaining patterns of conflict between Burger and his colleagues).

43. Over his twelve years as an active status judge, Burger averaged 10.8 separate opinions per year, compared to the D.C. Circuit average of 9.51 separate opinions per “judge year.” During his ten years of active status duty, Blackmun averaged 1.8 separate opinions a year, as opposed to the Eighth Circuit average of 1.34. Burger’s comparative record is perhaps more impressive because he wrote more majorities than did Blackmun, relative to circuit norms, and thus had somewhat fewer opportunities to write separately. See infra app., tbl.3.

Although a standardized score was used for majority opinions, see infra app., tbl.1, such a measure is not as applicable for dissents and concurrences. When implementing a standardized score, one assumes the observed outcomes are generated by a normal distribution. In the Eighth Circuit, however, judges often did not write a separate opinion for the entire year. A substantial part of the estimated distribution would therefore exist below zero, and the censoring of observations at zero would result in bias in the observed mean and standard deviation. A standardized score would not have this problem for the D.C. Circuit based on annual outputs of separate opinions, but the utility of this score would be limited inasmuch as circuit comparisons cannot be made. Accordingly, a simple average score was used in Table 1, see infra app. See generally Moore, supra note 15, at 248-55 (discussing relationship between standard scores and normal distributions).
cuit court colleagues who are appointed by a President of the same political party. Measuring partisan autonomy was also a challenge in this setting; I did not attempt to gather comprehensive data on panel compositions by party for every judge on the D.C. and Eighth Circuits. While the record on political autonomy is thus less than complete, the evidence I did assemble on Judges Burger and Blackmun illustrates some of the difficulties involved in the Choi and Gulati approach to measuring such autonomy.

During Judge Burger’s years on the D.C. Circuit, most of his colleagues were Democratic appointees: among his seven to eight fellow active status judges, there were never more than two appointed by Republican Presidents. It is therefore not terribly surprising that sixty-eight of his seventy dissents were from majority opinions written by Democratic appointees; for only eleven of these seventy dissents was there even a “panel majority” of Republicans, meaning a second Republican besides Burger himself.

Judge Blackmun belonged to a more politically balanced court of appeals. In his ten full years on the Eighth Circuit, Blackmun served mostly with Republican appointees during his early years and mostly with Democratic appointees in his later period. The less partisan context of Blackmun’s eleven dissents reflects this relatively balanced circuit court composition: four dissents were from all-Republican majorities, two were from all-Democratic majorities, and the other five were from “mixed” majorities of one Republican and one Democratic appointee. The fact that the Eighth Circuit had a less partisan “edge” may also help explain why there were notably fewer dissents or concurrences published in response to Blackmun majorities than was the case for majorities authored by Burger.

44. See Choi & Gulati, Tournament, supra note 1, at 310 n.29.
45. Of the eighty-nine active status “judge years” served by Burger’s colleagues, see infra app., tbl.3, sixty-nine (78%) were by Democratic appointees and only twenty (22%) were by Republican appointees. Panel composition was even more heavily Democratic given that five of the six D.C. Circuit judges who served on senior status were Democratic appointees. Thus, the odds of having an all-Republican or even majority-Republican panel were extraordinarily low.
46. In fifty-nine of seventy decisions, Burger dissented from a majority joined by either two Democrat-appointed panel members or (in nine en banc cases) a majority composed overwhelmingly of Democrat-appointed judges. In ten other decisions, the majority triggering Burger’s dissent consisted of one Democrat and one Republican appointee. The one dissent that Burger authored from an all-Republican majority was in Price Broadcasters, Inc. v. FCC, 295 F.2d 166 (D.C. Cir. 1961) (Burger, J., dissenting).
47. Of the fifty-seven active status “judge years” served by Blackmun’s colleagues, twenty-eight (49.1%) were by Republican appointees and twenty-nine (50.9%) were by Democratic appointees. See infra app., tbl.3.
48. Judge Blackmun’s 210 majorities drew a total of five dissents and five concurrences from his active status Eighth Circuit colleagues. By contrast, Judge Burger’s 254 majorities elicited thirty-one concurrences and fifty-eight dissents by judges on the D.C. Circuit. Four of the five dissents to Blackmun came from Democratic appointees, fifty-four
A closer look at some details regarding the eleven Blackmun dissents suggests ways in which the partisan classification scheme becomes problematic. One Republican appointee from whose majority opinions Blackmun dissented on four occasions was appointed to the federal trial bench by President Roosevelt and elevated to the appellate court by President Eisenhower after thirteen years as a Democratic appointee.49 There is some question as to whether this judge’s partisan label should differ from that of a “pure” Republican appointee. Moreover, in at least two of his four dissents from all-Republican majority decisions, Blackmun’s opinion disagrees with the “liberal” position adopted by his Republican colleagues.50 When Choi and Gulati place special value on a judge’s autonomy in being willing to risk the displeasure of like-minded colleagues, they would seem to have in mind the judge’s departure from a shared partisan-related ideology, not its reinforcement. Accordingly, Blackmun’s record of dissents as evidencing political autonomy is at best ambiguous.

Turning to a less quantitative perspective, Judge Burger’s propensity to express himself through dissenting opinions was viewed by the political branches as a sign of positive Supreme Court potential. At his confirmation hearing, Senators approvingly referred to and quoted from five separate Burger dissents that had addressed controversies involving matters of criminal procedure, mental health

49. Judge Vogel is the judge referred to in text. He authored three majorities from which Blackmun dissented and was listed as lead judge on a per curiam majority that elicited a Blackmun dissent. Two of these four majorities were “pure” Republican and the other two were mixed majorities of a Republican and a Democrat. 50. See Bookwalter v. Phelps, 325 F.2d 186, 189-91 (8th Cir. 1963) (Blackmun, J., dissenting) (dissenting from decision affirming district court judgment in estate tax case that widow’s allowance qualified for marital deduction; Blackmun would have reconstrued applicable state law to take deduction away from the surviving spouse); Kroger Co. v. Doane, 280 F.2d 1, 6-7 (8th Cir. 1960) (Blackmun, J., dissenting) (dissenting from decision that affirmed jury verdict favoring plaintiff in diversity action for negligence; Blackmun contended it was reversible error not to submit defendant’s jury instruction on contributory negligence). The two other cases in which Blackmun dissented from majorities joined by two Republican appointees are more difficult to categorize on a liberal-conservative spectrum. See Wm. F. Crome & Co. v. Vendo Co., 299 F.2d 852, 852-53 (8th Cir. 1962) (Blackmun, J., dissenting) (dissenting from decision affirming district court judgment of patent infringement; Blackmun viewed evidence of inventiveness as inadequate to warrant patent protection); United States v. Wiley’s Cove Ranch, 295 F.2d 436, 446-52 (8th Cir. 1961) (Blackmun, J., dissenting) (dissenting from decision affirming district court holding that payments for livestock feed under Emergency Feed Program were properly certified by a county committee of Farmers’ Home Administration; Blackmun would have allowed for more searching review by Department of Agriculture of circumstances surrounding county committee’s certification).
law, and school desegregation.\footnote{Taking their cue from President Nixon's stated intention to appoint “strict constructionists” to the Supreme Court,\footnote{Nomination of Warren E. Burger, of Virginia, to Be Chief Justice of the United States: Hearing Before the S. Comm. on the Judiciary, 91st Cong. 5-6 (1969) [hereinafter Burger Hearing] (remarks of Sen. McClellan, referring with approval to Burger dissents in \textit{Killough v. United States}, 315 F.2d 241 (D.C. Cir. 1962) (en banc), and \textit{Frazier v. United States}, 419 F.2d 1161 (D.C. Cir. 1969); \textit{id.} at 7 (remarks of Sen. Ervin, referring with approval to Burger dissent in \textit{Frazier}); \textit{id.} at 15-16 (remarks of Sen. H. Byrd, referring with approval to Burger dissents in \textit{Frazier, Kent v. United States}, 401 F.2d 408 (D.C. Cir. 1968), \textit{Scott v. Macy}, 402 F.2d 644 (D.C. Cir. 1968), and \textit{Smuck v. Hobson}, 408 F.2d 175 (D.C. Cir. 1969) (en banc)). By contrast, the Senators questioning Burger referred to only one of his majority opinions. See \textit{id.} at 7 (remarks of Sen. Ervin, referring with approval to Burger majority in \textit{Powell v. McCormack}, 395 F.2d 577 (D.C. Cir. 1968), \textit{aff'd in part, rev'd in part}, 395 U.S. 486 (1969)).} several Senators praised Burger for dissents that suggested to them a willingness to help rein in Warren Court activism in areas of criminal law and separation of powers.\footnote{Id. at 3 (statement of Sen. H. Byrd, referring to Burger's judicial record as consistent with the President's philosophical expectations for Supreme Court appointees); see supra note 4.}

Given the strongly Democratic complexion of the appellate court on which Burger served, the fact that his disagreements were almost always with Democratic appointees, while virtually inevitable, may well not satisfy the Choi and Gulati autonomy measure for judges taking an unbiased approach to individual cases.\footnote{Burger Hearing, supra note 51, at 3 (statement of Sen. H. Byrd); \textit{id.} at 7 (remarks of Sen. Ervin); \textit{id.} at 15-16 (remarks of Sen. H. Byrd); see also 115 \textit{Cong. Rec.}, 15,176 (1969) (remarks of Sen. Dirksen, praising Judge Burger’s “courage and conviction” as evidenced in his “many concurring and dissenting opinions”); \textit{id.} at 15,179 (remarks of Sen. Holland, identifying with Judge Burger’s philosophy as “firmly worded and firmly expressed” in his dissenting opinions). Whether these dissents were written in part to campaign for a future seat on the Supreme Court is a question beyond the scope of this Essay, although the Choi and Gulati approach might well encourage such strategic judicial behavior. See generally Julius Duscha, \textit{Chief Justice Burger Asks: 'If It Doesn’t Make Good Sense, How Can It Make Good Law?'}, \textit{N.Y. Times}, Oct. 5, 1969, §6 (Magazine), at 30 (reporting that in 1967, a mutual friend of Richard Nixon and Warren Burger told Burger of Nixon’s favorable reaction to a Burger speech criticizing two Warren Court decisions that had expanded the rights of criminal defendants).} Still, the frequency and clangor of these disagreements do suggest that Burger had the courage of his convictions in a largely hostile doctrinal environment, as well as a determination to exert extra effort to voice those convictions—traits that register as positives on the Choi and Gulati scale.\footnote{See Choi & Gulati, \textit{Tournament}, supra note 1, at 310 n.29 (focusing on dissents from politically like-minded judges as indicating lack of bias).}

This abbreviated comparison between two appellate judges is hardly meant to be definitive. I have not evaluated Judges Burger or Blackmun in the more detailed context of how appellate judges from all circuits conducted themselves individually during the same time

\footnote{See Choi & Gulati, \textit{Empirical Ranking}, supra note 1, at 62 (extolling extra effort and willingness to displease colleagues).}
period. Overall, though, Judge Burger emerges as a candidate of somewhat greater potential under the Choi and Gulati approach. Burger’s productivity in publishing majority opinions exceeded Blackmun’s. His majority opinions were cited by other circuits at the same rate as Blackmun’s. It is true that Blackmun’s citation numbers are high for his Circuit, but Burger’s favorable average compared to his nationally respected circuit court peers stands out even more. Further, Burger’s persistent authorship of dissents and concurrences would more clearly establish his capacity for independence under the Choi and Gulati framework.

B. The Supreme Court Performance of Justices Burger and Blackmun

In considering the Supreme Court careers of these two long-serving Justices, I focus on one aspect of each man’s record, an aspect that suggests how difficult it is to predict future Supreme Court behavior based on the Choi and Gulati appellate court performance criteria. For Burger, the independence factor turns out to be normatively troubling. Judge Burger’s appellate court record of authoring numerous dissents and concurrences appears in retrospect to have signaled an aloofness from colleagues and a lack of consensus-building capability amply demonstrated during his tenure as Chief Justice. For Blackmun, the fact that he changed substantially in doctrinal and ideological terms while a member of the Supreme Court is closely linked to the recognition he has received as a Justice. Blackmun’s evolution, however, seems wholly unrelated to his measurable outputs as a member of the Eighth Circuit.

1. Justice Burger and Separate Opinions

I have noted Judge Burger’s propensity for publishing dissents and concurrences while serving on the D.C. Circuit; in addition to exceeding the “independent” expression of almost all his circuit colleagues, his record of dissents attracted the approving attention of the executive and legislative branches. With hindsight, however,

56. I do not compare the two Justices’ Supreme Court records based on the Choi and Gulati performance factors. While it would be possible to do so, such comparisons inevitably would raise additional questions. For instance, Burger published more majority opinions than Blackmun during the sixteen Terms they served together (243 to 219). See Table I(A) at the back of issue one (the annual Supreme Court review issue) of the Harvard Law Review, volumes 85 through 100. One could debate whether increases in individual productivity are more of a virtue on appellate courts that frequently confront caseload backlogs than on a Supreme Court that necessarily clears its calendar every year. Even assuming, however, that the modest differential between the two Justices (15.2 versus 13.7 per Term) constitutes a “productivity advantage” for Burger, the advantage may well be due to reasons unrelated to merit, such as Burger’s control over opinion assignments and the increasingly strained professional and personal relations between the two men.
Judge Burger’s inclination to write separate opinions seems more reflective of a standoffish and at times insensitive judicial style than of the unbiased deliberative approach anticipated under the Choi and Gulati framework. From a doctrinal standpoint, Burger remained reasonably consistent in his wary stance toward the rights of criminal defendants, but his record is more fitful or uneven in certain other areas of the law. More broadly, the Burger Court has been criticized for a legacy of “rootless activism” and the absence of an identifiable agenda or set of values. It may be unfair to expect that any Chief Justice could have imposed a coherent philosophy or direction on the Court during such a divisive period of intellectual and political ferment in the larger society. Nonetheless, Burger’s “independent” judicial approach and style were not conducive to promoting collegiality or building consensus.

By viewing a greater volume of separate appellate court opinions as a fundamentally positive indicator, Choi and Gulati expect that a judge’s willingness to express her views independently will enhance the objectivity of judicial decisionmaking. Although this judicial willingness may reflect a neutral sensibility and intellectual fortitude that can improve the quality of the deliberative enterprise, a judicial appetite for independent expression may additionally or alternatively promote disharmony and lack of cohesion on an appellate court. The prospects for divisiveness may increase if the judge’s insistence on the importance of separate expression is carried over to his new role as Chief Justice on the Supreme Court—where the controversies are closer, the stakes are higher, and judges participate continuously with the same colleagues in the dynamic of shared decisionmaking.

Burger himself seems not to have anticipated the magnitude of his transition to the Supreme Court. As he told the Senate Judiciary Committee, “I would conceive my judicial duties to be essentially the same, basically the same as they have been as a member of the U.S. court of appeals—deciding cases.” Consistent with that vision, Bur-

60. See Alschuler, supra note 58, at 1454; Nichol, supra note 59, at 323-24.
61. See Choi & Gulati, Empirical Ranking, supra note 1, at 62.
ger as Chief Justice continued to voice his independence from his judicial colleagues. He wrote far more dissents as a percentage of his majorities than any Chief Justice in the modern era and similarly authored substantially more concurrences per majority than other Chief Justices over the past eighty years.\(^{63}\) This appetite for writing separately contributed to Burger's reputation as a less than successful leader of the Court; it is of a piece with descriptions of imperious or tactless approaches toward his colleagues that impeded the crafting of consensus on a range of complex and hotly contested issues.\(^{64}\)

At the appellate court level, judges are often critical of any pronounced tendency to write separate opinions. They fear the erosion of institutional integrity that may result from a regular insistence on voicing one's own doctrinal positions when compromise or acquiescence would yield unanimity.\(^{65}\) They also worry that a judicial inclination to write separately may reflect a somewhat arrogant unwillingness to deliberate and genuinely consider alternative views, an unwillingness that in turn leads to poorer work products.\(^{66}\) On the other hand, judges who regularly publish separate opinions are not all regarded with disfavor. There are Justices whose record of authoring separate opinions is praised as a mark of commitment to unbiased principles\(^{67}\) or to a consistent vision of the law.\(^{68}\)

---

63. See Zane, supra note 58, at 977-78, 1006-07 (reporting on number of opinions authored by Chief Justices since 1921 Term). As a result of his persistent inclination to issue large numbers of dissents and concurrences, Burger authored more separate opinions—dissents plus concurrences—than majorities in his seventeen years as Chief Justice, a record not remotely approached by the six other Chief Justices. Id. at 1006 tbl.I. Zane's data include opinions written accompanying orders of the Court, such as those denying certiorari or relating to Supreme Court applications. Id. at 1006 n.213. With respect to opinions filed in cases briefed and argued before the Court, Justice Burger's concurrences and dissents together are nearly half (47.2\%) his total output. See Table I(A) of Harvard Law Review's annual Supreme Court review issue, volumes 84 through 100.

64. Matthew Brelis, Court Improvements, Not Ideology, Called Main Legacy, BOSTON GLOBE, June 26, 1995, at 1 (reporting on perceptions of Burger's "pompous, almost regal attitude" that made it hard for him to build consensus); see also Mark Tushnet, Why the Supreme Court Overruled National League of Cities, 47 VAND. L. REV. 1623, 1630 (1994) (discussing Burger's clumsy collegial style).


68. See, e.g., Akhil Reed Amar, Hugo Black and the Hall of Fame, 53 ALA. L. REV. 1221, 1232-34, 1241 (2002) (discussing Justice Black's critical role in leading Court's movement over two decades to incorporate Bill of Rights into Fourteenth Amendment);
Whether an appellate judge’s insistence on independent expression signifies a future Justice unusually free of ideological predispositions or uncommonly insensitive to the needs of a collegial institution is a question worth asking. With hindsight, one can point to reservations expressed at the time of Justice Burger’s nomination, indicating that his independent-mindedness might well adversely affect his ability to lead the Court.69 The nature of those doubts suggests that at bottom, the relationship between separate opinion writing and laudable or lamentable qualities of judging is a contingent one. Insights into that relationship are more likely to be found by examining the judge’s individual biography, his personality and temperament, and the perception of his intellectual integrity among his peers than by measuring his past performance on an objective scale of production.

2. Justice Blackmun and Evolution on the High Court

While Blackmun’s appellate court record scores less well than Burger’s under the Choi and Gulati approach, the two men were viewed similarly in the course of the Supreme Court appointments process.70 During his confirmation hearing, Blackmun was characterized by Senators from across the political spectrum as a judge whose written opinions reflected a philosophy of judicial restraint and a clear respect for precedent.71 There are modest indications in Blackmun’s Eighth Circuit opinions and Senate testimony that he might behave on the Supreme Court in a manner less constrained than his

---

69. See, e.g., Duscha, supra note 53, at 30, 148 (reporting criticism from Burger’s ex-law partner (who had since become a state supreme court justice) that Burger’s stubbornness and tenacious adherence to his convictions helped make him an excellent lawyer but have been liabilities for him as an appellate judge); Sidney E. Zion, Nixon’s Nominee for the Post of Chief Justice: Warren Earl Burger, N.Y. TIMES, May 22, 1969, at 36 (reporting the “well known [view] in the legal community that [Burger’s] professional differences with a majority of his [D.C. Circuit] colleagues have often been so harsh as to create a mutual disrespect,” and that some of his colleagues regarded him as “unsuited by talent and temperament to lead the High Court”).

70. See Lee Epstein et al., The Role of Qualifications in the Confirmation of Nominees to the U.S. Supreme Court, 32 Fla. St. U. L. Rev. 1145, 1160, 1162 (2005) (reporting that Burger and Blackmun had identical “perceived ideology” scores during appointments process and that their “perceived qualifications” were very similar, with Blackmun’s score marginally higher).

71. Nomination of Harry A. Blackmun, of Minnesota, to Be Associate Justice of the Supreme Court of the United States: Hearing Before the S. Comm. on the Judiciary, 91st Cong. 2 (1970) [hereinafter Blackmun Hearing] (statement of Sen. Mondale); id. at 34 (remarks of Sen. Ervin); id. at 55 (remarks of Sen. H. Byrd); see also id. at 3 (statement of Sen. Mondale, referring to an editorial in the Rochester (Minn.) Post-Bulletin describing Blackmun as a “strict constructionist” of the Constitution . . . [and] a man of moderate, commonsense views”).
reputation suggested, but the general view at the time was that Blackmun would make few waves on the Court and would follow the lead of his long-time friend Warren Burger.

Of course, Justice Blackmun’s Supreme Court career departed dramatically from these expectations. Looking beyond his signature contributions in the areas of abortion and commercial speech, Justice Blackmun began by voting with Chief Justice Burger in nearly nine of ten decisions, but by the time Burger retired in 1986, the two were aligned only about one-half the time. Conversely, Justice Blackmun voted with Justices Brennan and Marshall in about 50% of all decisions during his first several years on the Court; by 1986, he was voting with each of them more than 80% of the time.

This shift toward more liberal voting patterns—which continued over his full twenty-four-Term tenure—reflects Blackmun’s changing views on a range of regularly contested issues. During his years on the Court, Blackmun became more willing to construe the Constitu-
tion broadly in order to protect individual civil liberties.\textsuperscript{79} He exhibited greater sympathy for persons living in conditions of economic hardship and emphasized the consequent importance of assuring access to legal channels for those less well-off.\textsuperscript{80} Blackmun notably modified his doctrinal positions with respect to the Court’s role in policing the lines between state and federal sovereignty,\textsuperscript{81} in deferring to the state to safeguard the interests of children,\textsuperscript{82} and in overseeing the implementation of capital punishment.\textsuperscript{83}

There has been no shortage of informed speculation regarding why Justice Blackmun’s performance differed so sharply from what was predicted for him. In terms of an evolution in judicial philosophy, scholars and commentators have referred to Blackmun’s heightened sensitivity to the plight of “outsiders” in our society, especially their need for and entitlement to safeguards within the legal culture.\textsuperscript{84} These observers have identified a concomitant erosion of Blackmun’s faith in government, particularly his belief in government’s ability and willingness to fulfill various protective functions for individuals and groups who are vulnerable or at risk.\textsuperscript{85} Blackmun’s philosophy did not simply emerge full-blown once he joined the Court, but the

\textsuperscript{79} Compare Cohen v. California, 403 U.S. 15, 27-28 (1971) (Blackmun, J., dissenting) (arguing that wearing a jacket that bore the words “Fuck the Draft” was an “absurd and immature antic” that was “mainly conduct and little speech”), with Bowers v. Hardwick, 478 U.S. 186, 208, 214 (1986) (Blackmun, J., dissenting) (contending that Court’s decision to allow prosecution for consensual homosexual activity in one’s own home betrayed privacy values deeply rooted in our constitutional traditions).

\textsuperscript{80} Compare Wyman v. James, 400 U.S. 309, 321-22 (1971) (rejecting as impertinent welfare recipient’s Fourth Amendment claim against a welfare-monitoring program), and United States v. Kras, 409 U.S. 434, 449 (1973) (rejecting indigent’s request to waive bankruptcy filing fee of fifty dollars and emphasizing that recommended installment payments would be “less than the price of a movie and little more than the cost of a pack or two of cigarettes”), with Beal v. Doe, 432 U.S. 438, 462-63 (1977) (Blackmun, J., dissenting) (criticizing decision to uphold congressional ban on use of Medicaid funds for nontherapeutic abortions as “condescending” and “punitive” toward indigent and financially helpless women), and Lassiter v. Dep’t of Soc. Servs., 452 U.S. 18, 46-47 (1981) (Blackmun, J., dissenting) (criticizing decision denying right to counsel for indigent woman facing proceeding to terminate her parental rights).


\textsuperscript{83} Compare Furman v. Georgia, 408 U.S. 238, 408-10 (1972) (Blackmun, J., dissenting) (agreeing that capital punishment as applied was not unconstitutional), with Herrera v. Collins, 506 U.S. 390, 430-32 (1993) (Blackmun, J., dissenting) (urging abandonment of death penalty as unconstitutionally arbitrary).


\textsuperscript{85} See Wasby, supra note 9, at 198; Note, supra note 84, at 722-25.
fact that it was rooted to some extent in various pre-Supreme Court experiences\textsuperscript{86} hardly explains what brought it to fruition.

One factor that seems to have contributed in important respects to the development of Justice Blackmun’s jurisprudence is simply his exposure to the crucible of Supreme Court service. Blackmun seemed continuously affected by the “awesome realization”\textsuperscript{87} of the Supreme Court’s final power to affect the lives of individuals and the relationships among institutions of government. That realization could have reinforced a faith in tradition or a deference to precedent that frequently accompanies appellate court service. Instead, it seemed to liberate Blackmun, making him more open to the Court’s role in addressing unforeseen problems and proposing new solutions.\textsuperscript{88}

In addition to his keen attention to the magnitude and novelty of the Court’s agenda, Blackmun may have been affected at a more personal level by a deterioration in his relationship with the Chief Justice. Looking back after a dozen years on the Court, Blackmun expressed frustration at the early public image of him as functioning under Burger’s control or influence.\textsuperscript{89} Further, Burger’s pattern of assigning majority opinions, apparently giving Blackmun an unusually small number of majorities in close or important cases and more than his share of the less glamorous Indian and tax decisions, may have played a role in the growing alienation between the two Justices.\textsuperscript{90}

In assessing Blackmun’s development as a Justice, Professor Pamela Karlan referred to his intense awareness of the “profoundly lonely business” of judging,\textsuperscript{91} suggesting that for a sensitive person

\textsuperscript{86} See supra note 72 and accompanying text; see also Blackmun Hearing, supra note 71, at 38-39 (displaying sensitivity to alienation of youth in modern society, referring to what he had learned from his own experience with his daughters); Note, supra note 84, at 723 (discussing Blackmun’s reliance on medical knowledge and historical context, gleaned in large part from experience as counsel to the Mayo Clinic, to deepen his understanding of the abortion issue while drafting Roe v. Wade).

\textsuperscript{87} Blackmun Hearings, supra note 71, at 43.

\textsuperscript{88} See Philippa Strum, Change and Continuity on the Supreme Court: Conversations with Justice Harry A. Blackmun, 34 U. RICH. L. REV. 285, 300 (2000); Note, supra note 84, at 734-36; cf. Blackmun Hearing, supra note 71, at 43 (responding to a question about the binding aspects of Court precedent by stating, “Judges, even Justices of the Supreme Court, are humans and I suppose attitudes change as we go along. . . . As times have changed, Justices have changed. People take a second look.”).


\textsuperscript{90} See Wasby, supra note 9, at 185, 196-97 (discussing a 1986 political science study showing that Blackmun ranked next to bottom in number of important cases he had been assigned and noting Burger’s penchant for giving Blackmun an unusually high number of unanimous or wide margin cases to write).

\textsuperscript{91} Karlan, supra note 10, at 185 (referring to Blackmun’s own statements on loneliness); see also Jenkins, supra note 89, at 61 (quoting Blackmun’s reference to Supreme Court service as “distinctly lonely”).
such as Blackmun, this loneliness may have “deepen[ed] the reservoirs of empathy.”

Observers also have pointed to Blackmun’s openness to change as a sign of maturity and a commitment to continued reflection; to his deep concern for fairness in judicial decisions, especially as it affected the proverbial “little guy”; and to his receptivity to modes of understanding complex human events that transcended legal knowledge or analysis.

Certain judicial attitudes associated with Justice Blackmun’s career are not without their detractors. A penchant for self-doubt and openness to change are viewed at times as hallmarks of timidity if not inconsistency. Blackmun’s emotionally descriptive attentiveness to the plight of society’s outsiders has been dismissed as overly sentimental and lacking rigor. Despite such critiques, Blackmun is held in high regard by a range of legal and political pundits, based on his capacity for growth while on the Court, his ability to blend careful craftsmanship with a strong sense of compassion, and his abiding awareness of how the law affects the circumstances and conditions of ordinary people.

92. Karlan, supra note 10, at 185.
96. See Garcia v. San Antonio Metro. Transit Auth., 469 U.S. 528, 557-59 (1985) (Powell, J., dissenting) (critical of Blackmun’s pivotal role in “precipitate overruling of multiple precedents” as undermining respect for Court’s authority); Jenkins, supra note 89, at 57 (discussing criticisms of Blackmun’s focus on resolving discrete disputes while failing to propound consistent theories of law). See generally Idleman, supra note 65, at 1392-93; Suzanna Sherry, Judges of Character, 38 WAKE FOREST L. REV. 793, 804 (2003). The tendency to evolve while a member of the Court may also be criticized as frustrating the legitimate role played by the politically accountable branches in selecting an ideologically suitable candidate. For discussion of the interaction between law and politics in the selection of Justices and in Supreme Court decisionmaking, see infra Part II.A.

Many of these observations were offered in the context of Justice Blackmun’s retirement from the Court or his death, moments when one would expect a laudatory tone. Yet there is a distinct contrast between what has been said about Blackmun and Burger on such occasions. Tributes to Justice Burger are notably less effusive; they tend to focus less on his defining doctrinal contributions or his judicial philosophy and more on his contributions to the administration of justice or his general respect for our history and traditions. See, e.g.,
From a historical standpoint, it is probably too early for a thorough evaluation of Justice Blackmun’s contributions. Some key qualities that make him a distinguished Justice to his supporters—empathy for the litigants before him, an abiding interest in fairness, and a receptivity to extralegal modes of analysis—are doubtless viewed as shortcomings by his critics. Whether one is an acolyte or a dissenter, however, the qualities that have tended to focus debate about Blackmun’s “merit” as a Justice are hardly reflected in his appellate court citation count or his pattern of dissents as a member of the Eighth Circuit. The Choi and Gulati approach is simply not relevant to this debate.

Justice Blackmun’s metamorphosis while a member of the Court may be more the exception than the rule. Still, the surprising nature of his arc on the Supreme Court is far from unique. Although some Justices have turned out more liberal or attentive to questions of redistribution than their pre-Court record would have led one to expect, there are others whose careers on the Court were distinctly more conservative or protective of the status quo than was anticipated at the time of their appointment.

Yet, insofar as predicting the future performance of Supreme Court candidates can be a hazardous business, reliance on quantitative indicia of appellate court outputs is unlikely to clarify the crystal ball. This is due in part to the very different nature of the dockets confronted by appellate judges and Supreme Court Justices. The Supreme Court’s discretionary caseload is determined by how shifting coalitions of interested colleagues react to emerging developments in constitutional advocacy and legislative policy, developments that over the long term are largely unforeseeable. For example, Justice


99. See Terri Jennings-Peletti, *In Defense of a Political Court* 113 (1999) (contending, based on a consensus among scholars, that Presidents have enjoyed a seventy-five percent success rate in predicting the future ideological course pursued by their Supreme Court appointees); id. at 119-21 (arguing that for most Justices who depart from ideological expectations, the departure may be attributed to a lack of relevant interest or attentiveness by the appointing President, and that Blackmun’s evolution while a member of the Court was unusual even among the twenty-five percent classified as surprises). But cf. Henry J. Abraham, *Justices, Presidents, and Senators: A History of the U.S. Supreme Court Appointments from Washington to Clinton* 50-52 (rev. ed. 1999) (quoting several eminent Court scholars and former Presidents emphasizing the President’s limited ability to predict the performance of Supreme Court nominees).

Blackmun’s role in shaping the Court’s positions on abortion and commercial speech—and the impact that role had on his overall performance as a Justice—could hardly have been anticipated based on his Eighth Circuit record addressing a mandatory caseload largely characterized by more mundane matters of statutory interpretation.

In addition, the impact of precedent is diminished, and the importance of collegial interaction increased, on a court of last resort in which nine individuals decide every case en banc. Given the role played by personal dynamics in this unusually intense repeat-player setting, a Justice’s lifespan on the Court becomes a further major, unpredictable factor. Blackmun’s career would surely look different if he had left the Court after ten years instead of twenty-four, if the Eisenhower appointee who retired in 1981 had been Brennan rather than Stewart, or if the replacements for Burger, Powell, and Marshall had been appointed by a Democratic President.

In the end, Supreme Court performance depends heavily on factors that are qualitative and personal. These factors include individual character and sensibilities, biographical experiences within and outside the law, the particulars of interaction with a subtly changing set of colleagues, the impact of a fluid and highly controversial docket, and length of tenure on the Court. The heavily subjective focus does not mean that predictions about performance will not continue to be a familiar aspect of the Supreme Court appointments process. As discussed in Part II, however, such predictions should be based on factors other than quantifiable production as an appellate judge. The Choi and Gulati approach would give Burger high marks for a performance criterion that turned out to represent one of his serious shortcomings as a Supreme Court Justice. Further, the Choi and Gulati focus on appellate court outputs fails altogether to account for the qualities that principally defined Blackmun’s Supreme Court tenure.

II. POLITICS AND NONQUANTITATIVE FACTORS

A. The Legitimate and Appropriate Role of Politics

A driving force behind the Choi and Gulati proposal to measure and rank judicial performance is the authors’ belief that “politics is primarily to blame” for our “abysmal” system of selecting Supreme

---

101. See Rao, supra note 97, at 34-35, 39-40 (discussing the impact of abortion cases on Justice Blackmun’s philosophy); William S. Dodge, Weighing the Listener’s Interests: Justice Blackmun’s Commercial Speech and Public Forum Opinions, 26 Hastings Const. L.Q. 165, 170-93 (1998) (discussing Blackmun’s role in shaping the Court’s position on commercial speech).

Court Justices. Acknowledging that the political branches must continue to play a formal role in the nomination and confirmation process, Choi and Gulati maintain that their tournament will at least require politicians to address merit-based considerations in more objective and transparent terms. Ultimately, though, the authors seek to supplant the existing politically-based system; they contend that their market-based approach is normatively preferable to the opaqueness and subjectivity that are endemic to the political model.

This effort to minimize if not eliminate the role of partisan and ideological considerations is in my view misguided. Initially, the Constitution in its design anticipates that politics will play an important part in the judicial selection process. One indicator of the extent of presidential and senatorial control is the Constitution’s silence regarding any minimum qualifications for the federal judiciary. Although minimum requirements are specified for the office of President and for members of Congress, the Framers entrusted the executive and legislative branches with complete discretion to determine judicial qualifications through their decisions regarding which individuals would be nominated and confirmed. From a historical standpoint, several factors may help explain the absence of objective criteria or threshold requirements for service on the Supreme Court. Whatever the explanation in original terms, there is an ongoing constitutional contemplation that the partisan preferences and ideological priorities of the politically accountable branches will play

103. Choi & Gulati, Tournament, supra note 1, at 301.
104. Choi & Gulati, Empirical Ranking, supra note 1, at 27, 29-30.
105. Id.; Choi & Gulati, Tournament, supra note 1, at 302-04.
106. See U.S. Const. art. II, § 1, cl. 5 (providing that President must be at least thirty-five years of age, a natural born citizen, and a fourteen-year resident of the United States); id. art. I, § 2, cl. 2 (providing that Representatives must be at least twenty-five years of age, citizens for seven or more years, and residents of their state); id. art. I, § 3, cl. 3 (providing that Senators must be at least thirty years of age, citizens for nine or more years, and residents of their state). Article III of the Constitution contains no comparable requirements for federal judges. See ABRAHAM, supra note 99, at 35 (noting the surprising absence of any constitutional requirements to become a federal judge).
107. See U.S. Const. art. II, § 2, cl. 2 (providing that the President has the power to appoint Supreme Court Justices “by and with the Advice and Consent of the Senate”). The recruitment and selection of federal judges through the political process contrasts with the civil service approach to judicial selection adopted in some European countries. See PERETTI, supra note 99, at 85.
108. See John R. Vile & Mario Perez-Reilly, The U.S. Constitution and Judicial Qualifications: A Curious Omission, 74 Judicature 198, 200-02 (1991) (suggesting that constitutional silence may be due to the absence of contemporary standards for legal education or training, making it difficult to specify a uniform “lawyer” qualification; the absence of minimum age or educational requirements for judges serving under state constitutions of the time; and a concern that restrictive qualifications would serve to intensify fears of an aristocratic judiciary).
a role in the selection process, “serv[ing] as effective majoritarian checks on the [Supreme Court’s] counter-majoritarian function.”

Such constitutional contemplation has become increasingly resonant in today’s legal and public policy circumstances. Constitutional and statutory interpretation are now regularly matters of intense political controversy, and both the executive and legislative branches have come to understand the importance of investing in the judicial selection enterprise.

It should not be surprising that the President and his agents regard Supreme Court decisionmaking as directly related to the real and perceived success of their policy agenda. During our prolonged period of divided government, as enactment of major legislative reforms has become a special challenge, the White House has paid more attention to court-centered strategies as a means of implementing changes in policy. Those strategies prominently include urging the Supreme Court to enforce certain legal rules expansively while arguing that the Court should act with restraint on other occasions involving related provisions of public law. Indeed, in managing embedded regulatory schemes that address politically contested subjects such as civil rights, workplace standards, and consumer health or safety, the executive branch has often altered its litigation approach depending on which political party is shaping the federal government’s Supreme Court agenda. Accordingly, the President’s in-

110. Arguments presented in the ensuing several paragraphs were initially developed in an earlier article. See James J. Brudney, Recalibrating Federal Judicial Independence, 64 OHIO ST. L.J. 149, 153-61 (2003).
111. Between 1968 and 2004, Congress (at least one chamber) and the Presidency were controlled by different parties seventy-five percent of the time; the exceptions were 1977-1980 plus 1993-1994 (all Democratic) and 2001 plus 2003-2004 (all Republican). Even in those exceptional times, the Senate has had between 41 and 49 members from the minority party, allowing for the reality or threat of a filibuster, except for the two year period from 1977-1978. See 2 CONGRESSIONAL QUARTERLY’S GUIDE TO U.S. ELECTIONS 1570 (John L. Moore et al. eds., 4th ed. 2001).
112. See Jeremy Rabkin, At the President’s Side: The Role of the White House Counsel in Constitutional Policy, LAW & CONTEMP. PROBS., Autumn 1993, at 63, 84-86 (discussing White House influence in shaping Justice Department strategy on civil rights litigation before the Supreme Court).
terest in appointing Supreme Court Justices based on political and ideological compatibility should be viewed as part of his effort to “integrate[e] the federal judiciary into the dominant lawmaking coalition.”

The Senate likewise has a considerable policy-related stake in the selection of Supreme Court Justices. Current majorities, and even filibuster-proof minorities, are strongly interested in confirming Justices who will not undermine preferred regulatory enactments or constitutional landmarks. More generally, because legislators expect courts in the future to be bound by the laws they enact in the present, they will be concerned not to endorse for a lifetime appointment any Justice perceived as unduly hostile to their current legislative agenda. A Congress worried about the potential for judicial defiance does remain free to monitor Supreme Court performance after the fact, at least with respect to high-profile regulatory statutes that legislators want to see vigorously applied. In practice, however, it tends to be both arduous and depleting for Congress to invalidate jus-


115. See, e.g., The Nomination of Robert H. Bork to Be Associate Justice of the Supreme Court of the United States: Hearings Before the S. Comm. on the Judiciary, 100th Cong. 467-68 (1987) (questioning of nominee by Sen. Metzenbaum with respect to enforcement of Occupational Safety and Health Act); The Nomination of David H. Souter to Be Associate Justice of the Supreme Court of the United States: Hearings Before the S. Comm. on the Judiciary, 101st Cong. 64-65 (1990) (questioning of nominee by Sen. Thurmond with respect to the validity of Congress’s efforts to limit number of post-trial appeals by death row inmates); id. at 53-55 (questioning of nominee by Sen. Biden with respect to foundation and scope of constitutional right to privacy).

116. See John Ferejohn & Barry Weingast, Limitation of Statutes: Strategic Statutory Interpretation, 80 GEO. L.J. 565, 581 (1992) (suggesting that each enacting Congress wants its laws enforced and sympathetically applied into the future, and that courts can encourage siting legislators to act carefully and deliberatively by interpreting earlier legislative products in a sensitive fashion); McNollgast, The Political Origins of the Administrative Procedure Act, 15 J.L. ECON. & ORG. 180, 185-86 (1999) (discussing risk that courts will try to impose their own policy preferences, subtly or profoundly altering a political compromise years after its enactment).

dicial interpretations of a federal statute with which most legislators disagree.\textsuperscript{118}

There are, of course, risks that undue emphasis on political or ideological background during the selection process may undermine the Court's basic decisionmaking function. Senators and Presidents have professed their awareness of this risk when they publicly eschew the use of litmus test screening that asks candidates to make advance commitments on specific issues of constitutional or statutory interpretation.\textsuperscript{119} At the same time, consideration of ideological background for its general predictive value would not appear to jeopardize the principled core of judicial decisionmaking, especially when—as is often the case—the White House and the Senate focus on that background in an effort to temper perceived excesses by the other branch.\textsuperscript{120} Moreover, when the Senate fails to review candidly a

\textsuperscript{118} The difficulties stem both from the lack of time and information needed to monitor statutory interpretation decisions and from the procedural and resource constraints that inhibit legislative success. \textit{See John W. Kingdon, Agendas, Alternatives, and Public Policies} 39-43, 194 (1984) (discussing limitations on members' access to sophisticated policy information and on the political capital available to each Senator or Representative); James J. Brudney, \textit{Congressional Commentary on Judicial Interpretations of Statutes: Idle Chatter or Telling Response?}, 93 Mich. L. Rev. 1, 21-26 (1994) (discussing finite resources and limited windows of opportunity that restrict Congress's legislative capacity); Stefanie A. Lindquist & David A. Yalof, \textit{Congressional Responses to Federal Circuit Court Decisions}, 85 Judicature 61, 63-64, 68 (2001) (reviewing 966 committee reports accompanying every bill reported out of House, Senate, or conference committee from 1990 to 1998, and finding that enacted bills responded to 65 circuit court cases (clarifying, codifying, or overriding) and that reports referred to a total of 187 specific circuit court cases out of more than 200,000 decisions in that nine year period).

\textsuperscript{119} \textit{See, e.g.}, \textit{Judicial Nominations 2001: Should Ideology Matter?: Hearing Before the S. Subcomm. on Admin. Oversight and the Courts of the S. Comm. on the Judiciary, 107th Cong. 30} (2001) (statement of Sen. Hatch, contending that "the Senate's responsibility to provide advice and consent [should] not include an ideological litmus test, because a nominee's personal opinions are largely irrelevant so long as the nominee can set those opinions aside and follow the law fairly and impartially as a judge"); Nat Hentoff, \textit{To Get a Supreme Court Seat}, Wash. Post, Aug. 14, 1999, at A17 (reporting then-presidential candidate George W. Bush's statement that he would not require an ideological litmus test for the Supreme Court). \textit{But cf. id.} (reporting then-presidential candidate Bill Clinton's 1992 statement to Bill Moyers that he would want his first Supreme Court appointee to be a strong supporter of Roe v. Wade, although it made him uncomfortable to be taking such a litmus test position).

nominee’s ideology or judicial philosophy, it tends instead to pursue alternative strategies that may be disingenuous if not unseemly.121

None of this is meant to suggest that political factors should be the primary qualification for ascending to the Supreme Court. There is an expectation that candidates should be exceptionally accomplished in terms of their professional abilities, temperament, and integrity. The American Bar Association has evaluated Supreme Court candidates on such merit-based grounds since the Eisenhower Administration; until recently, both Congress and the executive branch have utilized those evaluations.122

In addition, a focus on competence and integrity as essential elements in the appointments process is eminently reasonable from the President’s standpoint. As a regular repeat player in Supreme Court litigation, the executive branch should prefer Justices who are likely to apply language, precedent, and logical reasoning in largely rigorous fashion when deciding cases. The White House also may see some political value in appointing a “higher quality” Justice who is well received by the organized bar and the informed media, and may perceive a corresponding political cost in installing mediocre or disreputable individuals on the Court.123

---

121. These strategies have often included review of a nominee’s past non-ideological indiscretions or his asserted ethical misconduct in the private sphere. See, e.g., Adell Crowe, People Watch, USA TODAY, Sept. 29, 1987, at 4A (describing how a local Washington, D.C., paper obtained a list of videos rented by Supreme Court nominee Robert Bork’s wife to investigate whether Bork might have watched X-rated films); Steven V. Roberts, Ginsburg Withdraws Name as Supreme Court Nominee, Citing Marijuana ‘Clamor,’ N.Y. TIMES, Nov. 8, 1987, at A1 (describing how reports of Douglas Ginsburg’s marijuana smoking while a Harvard law professor led to his withdrawing as Supreme Court nominee); ROGER K. NEWMAN, HUGO BLACK: A BIOGRAPHY 239, 247-63 (2d ed. 1997) (discussing concerns fueled if not inspired by conservative anti-New Deal media and interest groups to reveal Senator Black’s earlier membership in the Ku Klux Klan as a means of defeating his Supreme Court appointment).


123. See SHELDON GOLDMAN, PICKING FEDERAL JUDGES: LOWER COURT SELECTION FROM ROOSEVELT THROUGH REAGAN 4 note c (1997) (discussing why it is both good presidential policy and good politics to recruit and nominate highly qualified judges). The unsuccessful effort by certain Senators to make a virtue out of mediocrity with respect to a Supreme Court nominee illustrates the outlier nature of such anti-merit sentiments. See ABRAHAM, supra note 99, at 11 (counting comments by Senator Hruska, floor manager for the Carswell nomination, that “[e]ven if he is mediocre there are a lot of mediocre
There will be an ample number of suitably accomplished individuals available for each open seat on the Supreme Court, even assuming a limited pool of well qualified candidates. In this setting, considerations of “merit” are a necessary and important element of the appointment process, but they cannot fully define that process. This is because, while the Supreme Court is expected to act in a principled and intellectually coherent manner when resolving disputes of law, it has long been understood that the Court’s legal propositions are replete with judgments and choices that have significant policy consequences. In appellate controversies involving two thoroughly briefed positions, norms of legal craftsmanship constrain the judges’ policymaking discretion. At the same time, these norms can often be met regardless of which side prevails; indeed, a Justice who wants to see his policy preferences implemented will presumably make every effort to provide a coherent, well-reasoned, and carefully crafted legal opinion.

Choi and Gulati’s tournament attaches little or no value to the inevitable interplay of law and policy in Supreme Court decisionmaking. By contrast, the real world sees considerable value in this interplay, as evidenced by the intense recognition it receives during the appointments process. The President and the Senate regularly seek to shade the future judicial philosophy of the Supreme Court toward their own broadly conceived policy preferences. The fact that their reliance on political and ideological considerations is susceptible to occasional misuse does not impeach the legitimacy of such reliance, judges and people and lawyers. They are entitled to a little representation, aren’t they, and a little chance? We can’t have all Brandeises, Cardozos, and Frankfurters, and stuff like that there” (internal quotation marks omitted). But cf. John W. Dean, The Rehnquist Choice: The Untold Story of the Nixon Appointment That Redefined the Supreme Court 125-26, 133-42, 187 (2001) (discussing President Nixon’s willingness to consider mediocre or even outrageous candidates for the Supreme Court in 1971, at least partly as a form of payback for the Senate’s previous rejection of his nominees).

124. For reasons discussed in Part I, supra, it is far from clear that Choi and Gulati’s measurable performance factors based on appellate court outputs are preferable to the current, less quantitative approach as a way of evaluating the professional talents and temperament of prospective Supreme Court candidates.

125. See, e.g., Charles L. Black, Jr., The People and the Court: Judicial Review in a Democracy 173-74 (1960) (discussing widespread recognition that constitutional law consists of both “a great deal of sheer legal technicality” and profoundly value-laden judgments regarding public policy; because these questions of law and policy overlap and are inseparable, “[l]egal acumen [is] not a sufficient condition . . . for dealing competently with questions of constitutional law”); Karl N. Llewellyn, The Common Law Tradition: Deciding Appeals 374 (1960) (discussing the importance of reading a statute in light of some assumed purpose known to the legislature and observing that as a statute ages, an appellate court must apply it to circumstances unanticipated at the time of enactment, in effect extrapolating from the initial purpose to make new policy).

126. See Peretti, supra note 99, at 82, 159 (arguing that when Justices vote in accordance with their political preferences, a well-crafted and intellectually rigorous decision is less likely to activate the political sanctions that are occasionally used or threatened against the Court).
provided it is combined with a threshold concern for sufficiently meritorious candidates. Thus, for both principled and practical reasons, the two branches constitutionally charged with recruiting and selecting members of the Court will continue to search for an appropriate composite of law-related expertise and policy-related wisdom.

B. Collegiality and Career Diversity

In focusing on performance criteria that lend themselves to quantitative and comparative measurement, Choi and Gulati omit consideration of other, more “qualitative” factors that make critical contributions to the appellate decisionmaking process. I will touch briefly on two important aspects of individual judicial performance that are not easily analyzed at the empirical level—collegiality and career diversity.

1. Collegiality

Judges and judicial scholars have identified various ways in which collegiality provides shape and direction to group decisionmaking on an appellate court. As a process matter, collegial interaction tends to sharpen and deepen a court’s reasoning. Judges in conference who carefully engage and evaluate the arguments and explanations offered by their colleagues are likely to create a better informed and intellectually more rigorous final product even if no one’s vote is changed during deliberations. In addition, an individual appellate

127. To be sure, invocation of these policy-related considerations does not always produce the expected results, as aptly illustrated by the Burger and Blackmun stories. Despite their pre-Court similarity in political and ideological background, they turned out to differ markedly as Justices on a range of policy and value-laden controversies that came before the Court. Such surprises, however, will not deter the President and the Senate from efforts to identify individuals who combine an acceptably high level of legal craftsmanship with a suitably worthwhile set of policy preferences. See generally PERETTI, supra note 99, at 80-86, 130-32.

Choi and Gulati would presumably prefer that Justices’ ideological orientations be randomly distributed based on whoever has been found most meritorious under their three criteria. There are at least two problems with this approach, both alluded to in the text. First, it thwarts input from the political branches, input anticipated under our constitutional design. Second, it underestimates the sophistication and ingenuity of these political branches in being able to package or manipulate appellate judges’ “objective” scores in the service of an ideological agenda.

judge whose personal style helps foster a background norm of cordial, courteous relations may well play a more important role on contentious issues for which he has managed to facilitate continuing conversation.129

Collegiality affects decisionmaking at a substantive level as well, again operating in at least two distinct ways. By encouraging judges to modify their personal predilections and soften their traditional advocacy-oriented approach, the “filter” of collegiality can enhance prospects for a consensus that “mitigates judges’ ideological preferences.”130 In this respect, collegiality functions to constrain the influence of a judge’s personal vision or outlook. At the same time, collegiality can also produce consequences that are more ideologically directional. Judges whose reasoning skills and doctrinal vision are augmented by an ability to cultivate warm personal relationships with colleagues may be especially successful in forging alliances that, over time, determine a series of outcomes giving shape to an entire area of law.131

Although an integral part of the appellate decisionmaking process, collegiality is not readily measurable in the output-dependent terms relied upon by Choi and Gulati. Judge Harry Edwards has suggested that as a “qualitative variable . . . involv[ing] mostly private personal interactions,” collegiality may not be measurable at all.132 The ongoing private exchanges to which Judge Edwards refers seem likely to have an especially pervasive and subtle influence at the Supreme Court, where decisionmakers do more than simply come together from geographically dispersed locations on a periodic basis in order to resolve particular cases. The Justices effectively live together in professional terms, continuously inhabiting the same intellectual space; they have been described by one veteran Court observer as “locked into intricate webs of interdependence where the impulse to speak in a personal voice must always be balanced

129. See Gerhardt, supra note 102, at 1613-14 (discussing John C. Jeffries’ biography of Justice Lewis F. Powell and noting Jeffries’ suggestion that Powell’s “ingrained courtesy and ability to listen” may have enhanced his pivotal position in many instances even if he did not seek such a position).

130. Edwards, supra note 19, at 1689.

131. See Gerhardt, supra note 102, at 1611-12 (discussing Roger K. Newman’s biography of Justice Hugo Black and noting Newman’s contention that Black’s friendly demeanor, civility, and personal political skills enabled him to exert far greater influence on the Court than his chief rival, Justice Frankfurter, over a period of more than twenty years); Richard A. Posner, A Tribute to Justice William J. Brennan, Jr., 104 HARV. L. REV. 13, 14 (1990) (discussing how Justice Brennan’s personal qualities made him “the supremely collegial Justice” and observing that his consequent skill in constructing coalitions sustained his influence in doctrinal terms even after the Court’s balance of power had shifted away from him).

132. Edwards, supra note 19, at 1656. “Regression analysis does not do well in capturing the nuances of human personalities and relationships . . . .” Id.
against the need to act collectively in order to be effective.” It is therefore not surprising that the collegiality factor often figures prominently as part of in-depth examinations into the strengths or shortcomings of individual Supreme Court Justices.

To take just one example, the conference at which all nine Justices exchange views following oral argument affords an opportunity for wielding influence. Justice Thurgood Marshall’s colleagues have remarked on how deeply he affected their thinking during conferences, in ways that would hardly be amenable to formal measurement yet seem quite relevant when considering the “quality” of a Supreme Court Justice. Moreover, given that collegiality involves complex elements of personal chemistry, it is not clear that one’s reputation and accomplishments on an appellate court will be replicated or even approximated with an entirely new group of judicial peers. In sum, while the Choi and Gulati framework should not be expected to capture all merit-related dimensions of judging, its exclusive focus on quantifiable criteria overlooks a substantial and influential judicial attribute.

134. See Gerhardt, supra note 102, at 1610-14 (discussing how biographers of Justice Black and Justice Powell analyzed their respective subject’s collegiality and reached different conclusions).
137. See Duscha, supra note 53 (describing Judge Burger’s quiet methods of persuasion among his D.C. Circuit colleagues and suggesting that Burger had some success in building coalitions on that court).
138. Other important judicial attributes, such as integrity and temperament, may also be difficult or impossible to assess in empirical terms, but that discussion is beyond the scope of this Essay. See Ramo & Cooper, supra note 122, at 102-03 (describing qualitative elements that go into evaluating a nominee’s integrity and judicial temperament).
2. Career Diversity

An important presumption underlying the Choi and Gulati framework is that Supreme Court Justices should be selected primarily if not exclusively from the pool of sitting appellate court judges. It is true that in recent decades, federal appellate experience has become almost mandatory for appointment to the Supreme Court: seven of the last eight Justices chosen, and ten of the last thirteen, were serving on the circuit courts when nominated. Choi and Gulati are inclined to view this increasingly standard practice as not just inevitable but appropriate. While they would allow the pool to be expanded under certain conditions, they maintain that data chronicling appellate judges’ performance are likely to be the best predictor of future conduct on the Supreme Court, because no other apprenticeship has job responsibilities that are as closely analogous.

One difficulty with this presumption is that in reinforcing the current status quo, it may well sacrifice prospects for improved institutional decisionmaking by depriving the Court of more diverse career perspectives. Justice Cardozo famously referred to the value of having a “balance of eccentricities” to generate more reliable and respected legal standards. The value of the experiential range of those eccentricities is diminished when almost all Justices have appellate court judging as their most recent and by implication most meaningful professional exposure.

There is considerable evidence in the legal and social science literature that career diversity is significantly associated with the voting patterns of appellate judges. Professors Lee Epstein, Jack Knight, and Andrew Martin recently surveyed twenty-two studies that had investigated linkages between career experience and judicial outcomes, concluding that nearly seventy percent had found some sort of relationship. Those findings, combined with the broader literature suggesting that diversity enhances the collective

---

139. The only exceptions, dating back to 1967, are Justices Powell, Rehnquist, and O'Connor. Justice O'Connor was a sitting state appellate judge.
140. See Choi & Gulati, Tournament, supra note 1, at 318-20.
141. BENJAMIN N. CARDozo, THE NATURE OF THE JUDICIAL PROCESS 177 (1921) (“The eccentricities of judges balance one another. . . . [O]ut of the attrition of diverse minds there is beaten something which has a constancy and uniformity and average value greater than its component elements.”).
142. See Lee Epstein et al., The Norm of Prior Judicial Experience and Its Consequences for Career Diversity on the U.S. Supreme Court, 91 CAL. L. REV. 903, 961-65 (2003) (listing twenty-two studies exploring linkages between judges’ prior occupations and their decisionmaking); Brudney, supra note 110, at 169-70 (reviewing studies that have demonstrated associations between judicial voting patterns and experience as a law professor, experience as a prosecutor, and experience in elected office).
143. Epstein et al., supra note 142, at 954.
decisionmaking enterprise,144 should make us wary of too readily accepting or encouraging the value of homogeneous formative experiences for Supreme Court service.

A useful illustration is the virtually complete loss of the perspective of those Justices who had formerly held federal or state elected office. Until Chief Justice Warren’s retirement in 1969, the Court for nearly fifty years included anywhere from two to five Justices who had spent substantial time serving as U.S. Senators or Representatives, state legislators, governors, or (in one instance) President.145 Of twenty-one Justices appointed between 1921 and 1953, seven had previously been elected to federal office and six to state office.146 In stark contrast, of twenty Justices appointed since 1953, none had previously held federal elected office, and only one (Justice O’Connor) had been elected to serve at the state legislative or executive level.

While there are risks to generalizing about relations between the branches of government, it seems safe to observe that Supreme Court attitudes and approaches toward Congress’s lawmaking processes and Congress’s final work products were more deferential from the 1940s to at least the mid 1970s than they have become in more recent times.147 Many complex factors and circumstances have doubtless contributed to these changes at the Court, and there are diver-


145. From 1945 to 1956, five of the nine Court seats were occupied by individuals who had held federal or state elected offices, including Hugo Black (ten years in U.S. Senate); Stanley Reed (four years in Kentucky legislature); Harold Burton (four years in U.S. Senate); a seat occupied consecutively by Frank Murphy (two years as Michigan Governor) and Sherman Minton (six years in U.S. Senate); and a seat occupied consecutively by Fred Vinson (fourteen years in U.S. House) and Earl Warren (ten years as California Governor). 2 DAVID G. SAVAGE, GUIDE TO THE U.S. SUPREME COURT 975-1023 (4th ed. 2004).

146. The seven with federal elective office experience included three with Senate experience, one with House experience, two with both Senate and House experience, and one who had served as President. The six who had state elective office experience included three former governors and three former state legislators (two of whom also went on to serve in the U.S. Senate). See id.

147. See, e.g., GEOFFREY R. STONE, LOUIS M. SEIDMAN, CASS R. SUNSTEIN & MARK V. TUSCHNET, CONSTITUTIONAL LAW 185-203 (4th ed. 2001) (discussing evolution of less deferential Commerce Clause doctrine); id. at 220-30 (discussing development of less deferential Court perspective on Congress’s power to enforce Section 5 of the Fourteenth Amendment); id. at 233-55 (discussing evolution in Court’s willingness to imply Tenth Amendment limits on Congress’s powers); WILLIAM N. ESKRIDGE, JR., PHILIP P. FRICKER & ELIZABETH GARRETT, CASES AND MATERIALS ON LEGISLATION: STATUTES AND THE CREATION OF PUBLIC POLICY 742-58 (3d ed. 2001) (discussing current, less deferential Court perspective on the probative value of legislative history or “legislative intent” in understanding statutory meaning).
gent normative views as to the value of such changes. One might reasonably wonder, though, whether the Court’s dramatically decreased personal familiarity with how Congress or legislators generally operate has played a role in the doctrinal shift, as well as whether a Court dominated by former federal appellate judges is not perhaps affecting the tenor and direction of decisions in other ways.

Choi and Gulati are aware of the diversity problem; they have forthrightly offered to accommodate alternative career paths based on development of suitable measurement techniques. Yet, a focus on quantifiable performance criteria is likely to miss the larger and subtler ways in which appellate judges are influenced by the distinctive professional perspectives of their colleagues. One can envision an approach that measures legislators’ effectiveness based on their percentage of missed floor votes; the number of bills sponsored, hearings chaired, or live floor speeches delivered; and perhaps even the number of enacted public laws for which they receive total or partial credit. It is questionable whether these or related measurements can adequately capture what makes prior experience in Congress or state legislatures a valuable “qualifier” for Supreme Court service.

**CONCLUSION**

Professors Choi and Gulati are surely correct to insist on the importance of merit or competence in the selection process for Supreme Court Justices. They also rightly argue for a less disingenuous approach to merit-based assessment on the part of both Congress and the White House. However, their evaluative focus on the transparency and objectivity of appellate judge track records would not be an improvement on our current, admittedly imperfect, system.

The assumption that prior judicial experience should be a principal determinant of Supreme Court potential has itself been challenged by scholars of the Court. One need not fully embrace such

---


150. See Felix Frankfurter, *The Supreme Court in the Mirror of Justices*, 105 U. PA. L. REV. 781, 795 (1957) (concluding, upon review of Justices’ careers over 167 years, that “[o]ne is entitled to say without qualification that the correlation between prior judicial experience and fitness for the functions of the Supreme Court is zero”); Gregory A. Caldeira, *In the Mirror of the Justices: Sources of Greatness on the Supreme Court*, 10 POL. BEHAV.
challenges in order to question the predictive value of this experience as measured in purely quantitative terms. Viewed in a comparative setting, the appellate court outputs of Warren Burger and Harry Blackmun are not overly enlightening with regard to the very different agenda and dynamics that confront Supreme Court Justices. In addition, by dismissing politics as a major flaw in the selection process, Choi and Gulati ignore the legitimate and appropriately constraining roles played by both political branches. Finally, there are certain more qualitative factors important to the Court’s effective performance as a lawmaking institution; while these may be difficult to assess on an individual candidate basis, they should not be overlooked when considering each candidate’s potential to achieve greatness or to enhance the success of the Court.

247, 258 (1988) (using multivariate model to conclude, inter alia, that contrary to much conventional wisdom, there is no association between judicial experience and eminence as a Supreme Court Justice).


**APPENDIX**

**TABLE 1 – PUBLISHED MAJORITIES**

**D.C. CIRCUIT 1957-1968**

<table>
<thead>
<tr>
<th>Judge</th>
<th>Full Years Active</th>
<th>No. of Majority Opinions</th>
<th>Average Productivity Score</th>
<th>Circuit Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edgerton</td>
<td>6</td>
<td>82</td>
<td>-1.11</td>
<td>13</td>
</tr>
<tr>
<td>Miller</td>
<td>7</td>
<td>95</td>
<td>-1.17</td>
<td>14</td>
</tr>
<tr>
<td>Prettyman</td>
<td>5</td>
<td>99</td>
<td>0.39</td>
<td>4</td>
</tr>
<tr>
<td>Bazelon</td>
<td>12</td>
<td>220</td>
<td>-0.09</td>
<td>8</td>
</tr>
<tr>
<td>Fahy</td>
<td>12</td>
<td>227</td>
<td>0.26</td>
<td>5</td>
</tr>
<tr>
<td>Washington</td>
<td>8</td>
<td>152</td>
<td>0.10</td>
<td>6</td>
</tr>
<tr>
<td>Dahaner</td>
<td>12</td>
<td>214</td>
<td>-0.07</td>
<td>7</td>
</tr>
<tr>
<td>Bastian</td>
<td>8</td>
<td>127</td>
<td>-0.66</td>
<td>10</td>
</tr>
<tr>
<td>Burger</td>
<td>12</td>
<td>254</td>
<td>0.63</td>
<td>2</td>
</tr>
<tr>
<td>Wright</td>
<td>6</td>
<td>100</td>
<td>-0.24</td>
<td>9</td>
</tr>
<tr>
<td>McGowan</td>
<td>5</td>
<td>104</td>
<td>0.51</td>
<td>3</td>
</tr>
<tr>
<td>Tamm</td>
<td>3</td>
<td>41</td>
<td>-0.77</td>
<td>11</td>
</tr>
<tr>
<td>Leventhal</td>
<td>3</td>
<td>87</td>
<td>1.85</td>
<td>1</td>
</tr>
<tr>
<td>Robinson</td>
<td>2</td>
<td>26</td>
<td>-0.96</td>
<td>12</td>
</tr>
</tbody>
</table>

**Eighth Circuit 1960-1969**

<table>
<thead>
<tr>
<th>Judge</th>
<th>Full Years Active</th>
<th>No. of Majority Opinions</th>
<th>Average Productivity Score</th>
<th>Circuit Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Woodrough</td>
<td>1</td>
<td>14</td>
<td>-0.44</td>
<td>9</td>
</tr>
<tr>
<td>Johnsen</td>
<td>5</td>
<td>31</td>
<td>-1.56</td>
<td>12</td>
</tr>
<tr>
<td>Vogel</td>
<td>8</td>
<td>166</td>
<td>0.18</td>
<td>4</td>
</tr>
<tr>
<td>Von Ooster-Hout</td>
<td>10</td>
<td>255</td>
<td>0.76</td>
<td>1</td>
</tr>
<tr>
<td>Matthies</td>
<td>10</td>
<td>220</td>
<td>0.30</td>
<td>3</td>
</tr>
<tr>
<td>Blackmun</td>
<td>10</td>
<td>210</td>
<td><strong>0.01</strong></td>
<td><strong>6</strong></td>
</tr>
<tr>
<td>Ridge</td>
<td>4</td>
<td>54</td>
<td>-0.52</td>
<td>10</td>
</tr>
<tr>
<td>Meheffy</td>
<td>6</td>
<td>136</td>
<td>-0.28</td>
<td>8</td>
</tr>
<tr>
<td>Gibson</td>
<td>6</td>
<td>120</td>
<td>0.07</td>
<td>5</td>
</tr>
<tr>
<td>Lay</td>
<td>3</td>
<td>69</td>
<td>-0.18</td>
<td>7</td>
</tr>
<tr>
<td>Heaney</td>
<td>3</td>
<td>76</td>
<td>0.38</td>
<td>2</td>
</tr>
<tr>
<td>Bright</td>
<td>1</td>
<td>24</td>
<td>-1.06</td>
<td>11</td>
</tr>
</tbody>
</table>
### Table 2 – Citations Outside Own Circuit

**D.C. Circuit Majority Opinions 1963-1966**

<table>
<thead>
<tr>
<th>Judge</th>
<th>Number of Majority Opinions</th>
<th>Outside Circuit Cites Through 5/31/69—Overall Mean</th>
<th>Outside Circuit Cites Through 5/31/69—Top 20 Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bazelon</td>
<td>86</td>
<td>3.71</td>
<td>10.40</td>
</tr>
<tr>
<td>Fahy</td>
<td>85</td>
<td>3.33</td>
<td>9.65</td>
</tr>
<tr>
<td>Danaher</td>
<td>66</td>
<td>2.03</td>
<td>5.70</td>
</tr>
<tr>
<td>Burger</td>
<td>67</td>
<td>6.09</td>
<td>16.00</td>
</tr>
<tr>
<td>Wright</td>
<td>65</td>
<td>4.88</td>
<td>11.10</td>
</tr>
</tbody>
</table>

**Eighth Circuit Majority Opinions 1963-1966**

<table>
<thead>
<tr>
<th>Judge</th>
<th>Number of Majority Opinions</th>
<th>Outside Circuit Cites Through 5/31/69—Overall Mean</th>
<th>Outside Circuit Cites Through 5/31/69—Top 20 Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vogel</td>
<td>90</td>
<td>3.41</td>
<td>9.85</td>
</tr>
<tr>
<td>Van Oosterhout</td>
<td>111</td>
<td>2.90</td>
<td>8.15</td>
</tr>
<tr>
<td>Matthies</td>
<td>93</td>
<td>2.98</td>
<td>7.60</td>
</tr>
<tr>
<td>Blackmun</td>
<td>95</td>
<td>5.66</td>
<td>16.50</td>
</tr>
</tbody>
</table>
### Table 3 – Dissents/Concurrences

**D.C. Circuit 1957-1968**

<table>
<thead>
<tr>
<th>Judge</th>
<th>Full Years Active</th>
<th>No. of Majority Opinions</th>
<th>No. of Concurring Opinions</th>
<th>No. of Dissenting Opinions</th>
<th>Separate Ops. per Full Year</th>
<th>Cir. Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edgerton</td>
<td>6</td>
<td>82</td>
<td>7</td>
<td>29</td>
<td>6.00</td>
<td>9</td>
</tr>
<tr>
<td>Miller</td>
<td>7</td>
<td>95</td>
<td>17</td>
<td>72</td>
<td>12.70</td>
<td>3</td>
</tr>
<tr>
<td>Prettyman</td>
<td>5</td>
<td>99</td>
<td>7</td>
<td>17</td>
<td>4.80</td>
<td>12</td>
</tr>
<tr>
<td>Bazelon</td>
<td>12</td>
<td>220</td>
<td>58</td>
<td>135</td>
<td>16.10</td>
<td>1</td>
</tr>
<tr>
<td>Fady</td>
<td>12</td>
<td>227</td>
<td>40</td>
<td>89</td>
<td>10.75</td>
<td>5</td>
</tr>
<tr>
<td>Washington</td>
<td>8</td>
<td>152</td>
<td>23</td>
<td>30</td>
<td>6.60</td>
<td>8</td>
</tr>
<tr>
<td>Danaher</td>
<td>12</td>
<td>214</td>
<td>28</td>
<td>69</td>
<td>8.10</td>
<td>7</td>
</tr>
<tr>
<td>Bastian</td>
<td>8</td>
<td>127</td>
<td>5</td>
<td>39</td>
<td>5.50</td>
<td>10</td>
</tr>
<tr>
<td>Burger</td>
<td>12</td>
<td>254</td>
<td>60</td>
<td>70</td>
<td>10.80</td>
<td>4</td>
</tr>
<tr>
<td>Wright</td>
<td>6</td>
<td>100</td>
<td>42</td>
<td>53</td>
<td>15.80</td>
<td>2</td>
</tr>
<tr>
<td>McGowan</td>
<td>5</td>
<td>104</td>
<td>18</td>
<td>8</td>
<td>5.20</td>
<td>11</td>
</tr>
<tr>
<td>Tamm</td>
<td>3</td>
<td>41</td>
<td>5</td>
<td>21</td>
<td>8.70</td>
<td>6</td>
</tr>
<tr>
<td>Leventhal</td>
<td>3</td>
<td>87</td>
<td>9</td>
<td>5</td>
<td>4.70</td>
<td>13</td>
</tr>
<tr>
<td>Robinson</td>
<td>2</td>
<td>26</td>
<td>4</td>
<td>1</td>
<td>2.50</td>
<td>14</td>
</tr>
<tr>
<td>Total</td>
<td>101</td>
<td>1828</td>
<td>323</td>
<td>638</td>
<td>9.51 (Mean)</td>
<td></td>
</tr>
</tbody>
</table>

### Eighth Circuit 1960-1969

<table>
<thead>
<tr>
<th>Judge</th>
<th>Full Years Active</th>
<th>No. of Majority Opinions</th>
<th>No. of Concurring Opinions</th>
<th>No. of Dissenting Opinions</th>
<th>Separate Ops. per Full Year</th>
<th>Cir. Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Woodrough</td>
<td>1</td>
<td>14</td>
<td>0</td>
<td>5</td>
<td>5.00</td>
<td>2</td>
</tr>
<tr>
<td>Johansen</td>
<td>5</td>
<td>31</td>
<td>1</td>
<td>2</td>
<td>0.60</td>
<td>10</td>
</tr>
<tr>
<td>Vogel</td>
<td>8</td>
<td>166</td>
<td>1</td>
<td>4</td>
<td>0.62</td>
<td>9</td>
</tr>
<tr>
<td>Van Oosterhout</td>
<td>10</td>
<td>255</td>
<td>6</td>
<td>10</td>
<td>1.60</td>
<td>6</td>
</tr>
<tr>
<td>Matthies</td>
<td>10</td>
<td>220</td>
<td>4</td>
<td>5</td>
<td>0.90</td>
<td>7</td>
</tr>
<tr>
<td>Blackmun</td>
<td>10</td>
<td>210</td>
<td>7</td>
<td>11</td>
<td>1.80</td>
<td>5</td>
</tr>
<tr>
<td>Ridge</td>
<td>4</td>
<td>54</td>
<td>0</td>
<td>3</td>
<td>0.75</td>
<td>8</td>
</tr>
<tr>
<td>Mehaffy</td>
<td>6</td>
<td>136</td>
<td>0</td>
<td>3</td>
<td>0.50</td>
<td>11</td>
</tr>
<tr>
<td>Gibson</td>
<td>6</td>
<td>120</td>
<td>0</td>
<td>3</td>
<td>0.50</td>
<td>11</td>
</tr>
<tr>
<td>Lay</td>
<td>3</td>
<td>69</td>
<td>7</td>
<td>10</td>
<td>5.70</td>
<td>1</td>
</tr>
<tr>
<td>Heaney</td>
<td>3</td>
<td>76</td>
<td>0</td>
<td>6</td>
<td>2.00</td>
<td>3</td>
</tr>
<tr>
<td>Bright</td>
<td>1</td>
<td>24</td>
<td>2</td>
<td>0</td>
<td>2.00</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>67</td>
<td>1375</td>
<td>28</td>
<td>62</td>
<td>1.34 (Mean)</td>
<td></td>
</tr>
</tbody>
</table>
The United States Constitution provides that the President has the power to appoint federal judges with the advice and consent of the Senate.1 The Constitution does not specify the criteria that the President should use in selecting judicial nominees or that the Senate should employ in reviewing them. In recent years, the process of nominating and confirming candidates for the federal bench, and especially the Supreme Court, has become increasingly political and contentious. Professors Choi and Gulati criticize the apparently growing role ideology plays in choosing and evaluating judicial nominees and propose a bold alternative.2 Their “Tournament of Judges” purportedly consists of a series of ideologically neutral measures that identify which appellate judges “merit” elevation to the Supreme Court.3 By restricting the choice of a nominee to the winner of the tournament, Professors Choi and Gulati hope to eliminate the role of ideology and the attendant partisan battling from the selection of Supreme Court Justices. Moreover, they claim that their market-based system for judicial selection would improve the quality of nominees.4 The current federal appellate bench, which is itself a product of the very system that Professors Choi and Gulati lament, should perhaps be grateful for their providing the equivalent of an HR manual for boosting each judge’s odds of promotion.5 But we are

---

3. See Choi & Gulati, Tournament, supra note 2, at 301-02.
4. See id.; Choi & Gulati, Empirical Ranking, supra note 2.
5. Whether any judge should want to seek appointment to the Court is beyond Choi and Gulati’s study and our comments. Judge Frank Easterbrook has recently commented that “any judge who claims not to fancy a position on that Court is a liar.” Howard Bashman, 20 Questions for Circuit Judge Frank H. Easterbrook of
convinced that evaluating judicial performance is not as easy as they suppose and that relying exclusively on the Tournament to select a Supreme Court nominee would not advance the rule of law.

As a preliminary comment, we applaud Professors Choi and Gulati for the purposes for which they undertake their study. Improving the quality of public discourse on the composition of the judiciary and increasing public awareness of the judiciary’s work and how it functions are laudable goals. Moreover, empirical measurements, particularly empirical comparisons of a nominee to her judicial peers, can play an important role in informing decisionmakers and the public on the relative merit of a nominee. We disagree, however, that empirical measurements should be the sole basis on which a nominee should be chosen. In the end, the real mettle of a potential nominee to the Court lies in her opinions and character. Reading opinions (much less discerning character), however, is time-consuming and, hence, costly. It is, moreover, an inexact science. For members of the public without legal training, comprehending the often complex legal analysis of a judicial opinion is prohibitive. The public has turned instead to a less costly means of evaluating judicial nominees—looking at a nominee’s positions on what Professors Choi and Gulati call the “hot button issues,” such as abortion, gun rights, affirmative action, and capital punishment, among others. Professors Choi and Gulati decry the hot-button approach as unduly narrowing the range of legal issues discussed and reducing the confirmation process “to quibbling over [a nominee’s] expected position on issues like affirmative action and abortion.” They propose supplanting, not merely supplementing, the current nomination and confirmation process with their system of rankings. Their rhetorically charged use of the word “tournament” implies that the highest-ranked judge has won or earned a position on the Court and that the President and Senate should be reduced to the ministerial roles of simply awarding the tournament winner her rightful place on the Court.

An evaluation of whether the current nomination and confirmation system should be replaced with Choi and Gulati’s tournament scheme requires comparing the costs and benefits of each and determining which offers society the greatest benefits net of costs. Natur-

7. See Choi & Gulati, Empirical Ranking, supra note 2, at 35-36.
8. See id. at 24, 34, 39.
9. Id. at 33.
10. See id. at 35-36.
rally, as advocates of the tournament, Professors Choi and Gulati lament the shortcomings of the current hot-button system because it fails to distinguish arguments about a nominee’s merit from arguments about the nominee’s ideology. A politician’s advocacy of, or opposition to, a particular nominee for purely “political” reasons appears unseemly, and politicians consequently attempt to mask their ideological arguments as arguments on the merits. Implicit in Professor Choi and Gulati’s justification for their tournament is that politicians have become so successful at conflating ideology and merit that the public can no longer distinguish the two. The current senatorial stalemate on many appellate nominees, which many view as a dress rehearsal for a confrontation over a Supreme Court nominee, is presumably one consequence.

In this Essay, we describe several potential shortcomings of using the tournament to select judicial nominees. Before we turn attention to these potential costs, some benefits of the hot-button approach are worth noting. First, hot-button issues are politically charged because they matter to the public. Other issues about which the public should be concerned—but is not—certainly exist. The public’s failure to recognize the import of these other issues is a shortcoming of political discourse generally and is not specific to judicial confirmations; arguments over economic, domestic, and foreign policies are full of nuance and yet are frequently reduced to sloganeering and sound bites. Second, these issues may be hot buttons partly because they do have some power to predict a nominee’s positions on other, more obscure legal issues. Third, if a Supreme Court nominee has written an appellate opinion on the hot-button issue, the opinion should prove highly informative of the nominee’s abilities. When a judge confronts a case raising a hot-button issue, she knows that the public (and in the event of a nomination to the Supreme Court, the White House and Senate) will scrutinize the opinion, and she faces a strong incentive to put forth her best possible efforts. An appellate opinion on a hot-button issue should therefore represent the author’s best judicial efforts and provide significant insights into her abilities as a legal analyst and stylist. In effect, each hot-button issue may constitute its own tournament of judges, the equivalent of a law review writing competition for judges.

The emphasis of this Essay is not that the current system lacks room for improvement—few would argue that it does not need improvement—but that the brave, new world of Supreme Court nominations proposed by Professors Choi and Gulati may not be as clearly

11. Id. at 34-36.
12. Choi & Gulati, Tournament, supra note 2, at 301.
superior to the current system as they suppose. In this Essay, we
discuss three concerns with using the Tournament of Judges as a ba-
sis for selecting a nominee to the Supreme Court. First, like other
contributors to this Symposium, we question whether the metrics
proposed by Professors Choi and Gulati appropriately measure the
performance of circuit judges. Second, even if the Choi/Gulati metrics
accurately capture judicial performance, the tournament itself may
create incentives that distort judicial behavior and erode the quality
of appellate judging. The criteria and the method by which a judge
may improve her standing are readily known, and thus we are con-
cerned that reducing judging to finite, measurable results would en-
courage judges to promote tournament criteria rather than adjudi-
cate individual cases: judges may “judge to the test.” Third, even if
the Choi/Gulati metric accurately measures the performance of cir-
cuit judges, winning the tournament may not predict success as a
Supreme Court Justice.

I. HOW ACCURATE ARE THE CHOI/GULATI MEASURES?

A. Productivity

Professors Choi and Gulati measure judicial performance along
three dimensions: (1) productivity, (2) quality, and (3) independ-
ence.14 For productivity, they tally the number of opinions each judge
published, because they assume that unpublished opinions “often in-
volve minimal effort (and a lower quality of reasoning).”15 However,
norms about when a case warrants an opinion differ across circuits.
Professors Choi and Gulati correct for these differing norms by add-
ing to each judge’s tally the difference between the average number
of opinions published in the judge’s circuit and the average published
in the most productive circuit, which, currently, is the Seventh Cir-
cuit.16 The appeal of this measure is that a judge who resolves more
cases in published opinions appears hardworking, and it may often
be the case. However, the measure is not without difficulties.

First, Professors Choi and Gulati’s assumption that unpublished
opinions involve a lower quality of reasoning is dubious. In the Ninth
Circuit, the test for whether to publish a decision is whether existing
precedent squarely controls the issues presented.17 A judge’s failure

14. Id. at 42.
15. Id. at 43.
16. Id. at 45.
17. See 9TH CIR. GEN. ORDERS 4.3.a (“A memorandum disposition cannot be cited as
precedent. Unlike an opinion for publication which is designed to clarify the law of the cir-
cuit, a memorandum disposition is designed only to provide the parties and the district
court with a concise explanation of this court’s decision. Because the parties and the dis-
trict court are aware of the facts, procedural events and applicable law underlying the dis-
pute, the disposition need recite only such information crucial to the result.”); see also 9TH
to see novel issues in a case may indicate deficiencies in the judge’s intellectual curiosity and legal acumen, or it may reflect the judge’s respect for existing precedent. Cases vary in the novelty of the legal questions they present and, hence, in the degree of guidance the parties and future litigants require. As the Ninth Circuit rules indicate, cases that lack novelty should not be published because they offer current and future parties little or no instruction beyond existing precedent. In addition, judicial resources are constrained, at the very least, by the judge’s time. The opportunity cost of the time that a judge spends preparing a full opinion on a case furnishing little instruction to the public is the time that she could spend on another case presenting unresolved questions of law. Professors Choi and Gulati implicitly assume that all cases offer the same amount of instruction and the same amount of social benefit because they believe that all cases deserve published opinions. In their tournament, this assumption becomes an incentive to maximize publication. If pursued, this objective would work a substantial reallocation of judicial labor to less pressing questions. 18 The tournament encourages a substitution that would decrease rather than increase the amount of social benefit that the appellate courts provide.

Second, Choi and Gulati’s focus on published opinions is not devoid of ideological content. It disfavors advocates of judicial restraint who may prefer fewer published pronouncements from the bench. A recent study argues that the ideological leaning of a three-judge panel correlates with the decision to publish. 19 The very purpose of

---

18. See Richard A. Posner, The Federal Courts: Challenge and Reform 168-69 (1996) (“Given the workload of the federal courts of appeals today, the realistic choice is not between limited publication, on the one hand, and, on the other, improving and then publishing all the opinions that are not published today; it is between preparing but not publishing opinions in many cases and preparing no opinions in those cases.”); Hart v. Massanari, 266 F.3d 1155, 1177 (9th Cir. 2001) (“[F]ew, if any, appellate courts have the resources to write precedential opinions in every case that comes before them.”); id. at 1179 (“Adding endlessly to the body of precedent . . . can lead to confusion and unnecessary conflict.”).  

19. See David S. Law, Strategic Judicial Lawmaking: Ideology, Publication, and Asylum Law in the Ninth Circuit, 73 U. Cin. L. Rev. 817 (2005) (arguing that an examination of only published opinions leads to the erroneous conclusion that panels dominated by judges appointed by Republican Presidents were as likely to rule in favor of asylum seekers as panels dominated by judges appointed by Democratic Presidents).
the tournament is to develop a measure of judicial merit free of ideology, but by examining only published opinions, the tournament retains ideological content. It favors judges who distinguish each case from existing precedent and thereby foster complexity in the law.

Third, cases vary widely in their difficulty, but not in the Choi/Gulati rankings. In the tournament, a judge who resolves many small cases presenting simple issues in short opinions will rank higher than a judge who resolves a few large cases presenting complex issues in long opinions. In reality, the latter judge may have exerted greater effort and possess superior judicial talents. Fourth, even if the Choi/Gulati metric reflected pure productivity, intra- and intercircuit comparisons are a troublesome complication. In this measure, a judge who publishes more opinions than the average judge on his court and who sits in a circuit where the average number of opinions published per judge is above that of other circuits will be ranked higher than a judge who publishes fewer opinions than the average judge on his court and who sits on a circuit where the average number of opinions published per judge is below that of other circuits. However, the middling cases are less clear-cut. Is a judge who publishes more opinions than the average judge on his court but who sits in a circuit where the average number of opinions published per judge was below the average of other circuits better or worse than a judge who publishes fewer opinions than the average judge on his court but who sits in a circuit where the average number of opinions published per judge was above the average of other circuits? It is not obvious whether an above-average judge on a below-average circuit is better than a below-average judge on an above-average circuit. Without justification, Professors Choi and Gulati favor the former. They state that “[a]ny differences among judges will, therefore, be determined solely by each judge’s standing relative to the other judges within her own circuit,”20 and their productivity measure thus favors judges who are stars on their circuits. If their measure applied to basketball, Scottie Pippin playing for the Portland Trail Blazers would fare better than Scottie Pippin playing for the Chicago Bulls. In their metric, the presence of other all-stars who raise the circuit’s average productivity make it harder for a star judge to shine. Notably, judges on the D.C. Circuit, widely considered an all-star circuit and the warm-up bench for the Supreme Court, fare poorly in the Choi/Gulati productivity metric.21

21. See id. at 76-77.
B. Quality

Given their view of the current system of confirmation, Professors Choi and Gulati use a measure of quality that is inconsistent with their assumptions about judicial behavior. The tournament assumes that the number of citations an opinion receives from courts outside its circuit reflects its quality. Yet Professors Choi and Gulati believe that many nominees are advanced on the basis of ideology rather than merit. The tournament does not consider the possibility that the citations may themselves be objects of ideological manipulation rather than indications of the quality of the legal analysis. For citations to reflect solely quality, judges would have to shed their ideological leanings once they reach the bench. If judges did so, an ideological President would have no reliably ideological judges to nominate to the Court. Were it so, the problem of ideologically driven nominations would vanish, and the tournament itself would be unnecessary.

Even if judges do not choose their citations on the basis of ideology, Professors Choi and Gulati’s measure of quality may actually reflect other characteristics. They acknowledge that judges differ in the initial probability that their opinions are read. The notoriety a judge enjoyed before joining the bench, or enjoyed while on the bench but before the observation period of this study, carries over onto the study’s observation window. A lawyer known to legal commentators is more likely to have her opinions read once she takes to the bench, and hence she has a greater probability of having her opinions cited than does a judge whose prior legal career was relatively obscure. Professors Choi and Gulati claim that “it is precisely this [pre-tournament] reputation for quality analysis (and the underlying ability behind such a reputation) that we hope to find in a Supreme Court nominee.” The criterion favors academics over practitioners because publishing more readily establishes reputations than does winning cases or structuring deals. This measure promotes, for example, a Scalia, Ginsburg, or Breyer over a Powell, Blackmun, or Souter. Moreover, the labor force of the judiciary gives erstwhile (or part-time) legal academics on the bench a further advantage. Law clerks who draft opinions are typically recent graduates of law schools, where they studied the casebooks and commentaries of these

---

22. See id. at 48-50.
23. See id. at 30-31.
25. See Choi & Gulati, Empirical Ranking, supra note 2, at 49. “Our methodology does not control for such pretournament inherent differences.” Id.
26. Id.
judges qua academics. Having previously relied on the instruction of these professors’ writings, law clerks are likely to look to the opinions of these familiar oracles for guidance. Furthermore, a judge’s opinions may receive attention for reasons unrelated to his or her reputation for quality. The events that generate interest in a judge’s opinions may or may not stem from the trenchancy of his legal reasoning on the bench.27

The tournament’s quality measure may also reflect speed in rendering a decision on a common question faced by all circuits. Witness the recent flurry of opinions following the Supreme Court’s decision in Blakely v. Washington.28 The Blakely court held that the criminal sentencing guidelines of Washington State violated the Sixth Amendment because they did not require that every fact raising a sentence beyond the statutory maximum be proven to a jury beyond a reasonable doubt or admitted to by the defendant.29 Immediately, commentators and litigants wondered if the logic of Blakely would apply to the federal sentencing guidelines. Twelve days after the Supreme Court issued Blakely, the Seventh Circuit heard oral argument in United States v. Booker,30 and three days later, the court issued a decision. The majority in Booker held that Blakely implied that sentence enhancements under the U.S. Sentencing Guidelines based on facts not determined by a jury beyond a reasonable doubt violated the Sixth Amendment.31 Twelve days after Booker, the Ninth Circuit decided United States v. Ameline,32 in which the majority largely agreed with the Seventh Circuit: “We join the Seventh Circuit in holding that there is no principled distinction between the Washington Sentencing Reform Act at issue in Blakely and the United States Sentencing Guidelines.”33

In the six weeks following the respective decisions, courts in other circuits cited Booker twenty-eight times but cited Ameline only thirteen times. Despite a time lag of less than two weeks, the earlier decision was cited twice as many times. The advantage in terms of subsequent citations to the first court to resolve an issue is significant. Each circuit court looks to its sister circuits in confronting legal issues, in part because another circuit’s analysis may offer guidance.

29. See id. at 2537-38.
30. 375 F.3d 508 (7th Cir. 2004), aff’d and remanded by 125 S. Ct. 738 (2005).
31. See id. at 510, 513, 515.
32. 376 F.3d 967 (9th Cir. 2004), amended and superseded on reh’g, 400 F.3d 646 (9th Cir. 2005), reh’g en banc granted, 401 F.3d 1007 (9th Cir. 2005).
33. Id. at 974 (footnote omitted).
In addition, a decision that conflicts, rather than conforms, with that of another circuit is more likely to be called en banc or to prompt a certiorari grant. The first circuit to weigh in on a legal issue is therefore more likely to be cited than subsequent circuits that agree with its conclusion. The tournament’s quality measure may reflect this first-mover advantage rather than which opinion contains the most thorough analysis.

The first answer is not necessarily the right one. Substantively, the circuits split on Blakely’s relevance to the federal guidelines, and after granting certiorari in United States v. Booker, the Supreme Court recently resolved the split. The legal community viewed several aspects of the Court’s holding as a “surprise,” namely, the conclusions that judges must consult with the guidelines but that they are only advisory, and that sentences are subject to reasonableness review on appeal. Neither the author of the Seventh Circuit Booker majority (Judge Posner) nor its dissenter (Judge Easterbrook) nor any other appellate judge chose the same remedy as the Supreme Court did. The failure of any circuit judge to anticipate the Supreme Court’s holding during the few months between the Blakely decision and the grant of certiorari in Booker illustrates that fast analysis is not always the same as the Supreme Court’s analysis.

Another consideration is that opinions with clever turns of phrase are cited more often. A well-known rule of appellate practice is that an appellant’s brief must contain the argument with the “appellant’s contentions” and the reasons for them, with citations to the authorities and parts of the record on which the appellant relies. One purpose of this rule is to economize on judicial resources by sparing judges the task of searching for litigant’s arguments. In United States v. Dunkel, the Seventh Circuit vividly expressed this rule: “Judges are not like pigs, hunting for truffles buried in briefs.” We found thirteen instances in which other circuit courts quoted this line in published opinions. We located another two unpublished opinions

---

34. 125 S. Ct. 738 (2005).
36. 927 F.2d 955 (7th Cir. 1991) (per curiam).
37. Id. at 956.
38. See Crossley v. Georgia-Pacific Corp., 355 F.3d 1112, 1114 (8th Cir. 2004); Malacara v. Garber, 353 F.3d 393, 405 (5th Cir. 2003); Coggin v. Longview Indep. Sch. Dist., 337 F.3d 459, 468-69 (5th Cir. 2003) (en banc) (Jones, J., dissenting) (noting “[t]he problem has been colorfully, if hyperbolically, described by our brethren on the Seventh Circuit”); Indep. Towers of Wash. v. Washington, 350 F.3d 925, 929 (9th Cir. 2003); Roska ex rel. Roska v. Peterson, 328 F.3d 1230, 1246 n.13 (10th Cir. 2003); Craven v. Univ. of Colo. Hosp. Auth., 260 F.3d 1218, 1226 (10th Cir. 2001); United States v. Stuckey, 255 F.3d 528, 531 (8th Cir. 2001); Downs v. L.A. Unified Sch. Dist., 228 F.3d 1003, 1007 n.1 (9th Cir. 2000) (paraphrasing); Gross v. Burggraf Constr. Co., 53 F.3d 1531, 1546 (10th Cir. 1995); Teague v. Bakker, 35 F.3d 978, 985 n.5 (4th Cir. 1994); Enplanar, Inc. v. Marsh, 11 F.3d
in other circuits that cite Dunkel for the proposition that cursory allegations are inadequate but do not quote the pigs-and-truffles language directly. An earlier case in the Seventh Circuit made the same point: Rule 28 implies that appellate courts are not responsible for scouring briefs and accompanying documents for poorly articulated and supported arguments. However, it expressed the point less colorfully, and other circuit courts have not cited it for this proposition. In United States v. Williams, a criminal defendant appealed his conviction in part on the ground that the trial court improperly excluded hearsay testimony, but the appellant failed to identify the portions of testimony he sought to have admitted at trial. The court held that the appellant’s conclusory allegation did not comply with Rule 28 and was therefore waived. It also noted, as it did in Dunkel but in less memorable language, that the court (and the prosecutor) are not responsible for locating the portions of the record relevant to the appellant’s claim: “Neither this court nor the United States Attorney has a duty to comb the record in order to discover possible errors.” We have found no published opinion in another circuit that cites this language or even cites Williams for this proposition. Clever turns of phrase can be helpful mnemonics or just plain amusing, but they do not necessarily indicate deeper legal reasoning. In fact, most legal writing texts instruct that workmanlike prose and the avoidance of novel turns of phrase best achieve the clarity valued in legal analysis. Although writing ability and legal reasoning are surely complementary skills, judges’ appetite for the one-liners of their colleagues implies that a few zingers may distort the tournament’s measure of “quality.”

1284, 1297 n.16 (5th Cir. 1994); Greenwood v. FAA, 28 F.3d 971, 977 (9th Cir. 1994); Sanjour v. EPA, 984 F.2d 434, 437 n.3 (D.C. Cir. 1993). 39. See El-Fadl v. Cent. Bank of Jordan, 75 F.3d 668, 675 (D.C. Cir. 1996); Larson v. Nutt, 34 F.3d 647, 648 (8th Cir. 1994) (per curiam). 40. 877 F.2d 516, 518 (7th Cir. 1989). 41. Id. at 518-19. 42. Id. at 519. 43. See BRYAN A. GARNER, THE ELEMENTS OF LEGAL STYLE 6-8 (2d ed. 2002) (emphasizing clarity); HENRY WEIHOVEN, LEGAL WRITING STYLE 61-82 (2d ed. 1980) (discussing the goal of simplicity). Although it is not generally recognized as a legal writing text, Dickens’ David Copperfield is instructive on the matter: ‘How do you like the law, Mr. Micawber?’ ‘My dear Copperfield,’ he replied. ‘To a man possessed of the higher imaginative powers, the objection to legal studies is the amount of detail which they involve. Even in our professional correspondence,’ said Mr. Micawber, glancing at some letters he was writing, ‘the mind is not at liberty to soar to any exalted form of expression. Still, it is a great pursuit. A great pursuit!’ CHARLES DICKENS, DAVID COPPERFIELD 524-25 (Modern Library Paperback ed. 2000) (1850).
C. Independence

Professors Choi and Gulati also attempt to measure whether a judge thinks independently rather than ideologically. They measure independence as the frequency with which a judge opposes a fellow panel member appointed by a President of the same political party, relative to the frequency with which the judge sits with judges of the same political party.44 The assumption of this measure is that an ideological judge is less likely to dissent from decisions joined by members of her own party, and an independent judge is as likely to dissent from a decision joined by members of her own party as she is to dissent from one joined by members of the other party. This model of judicial decisionmaking, that a judge’s political tendencies mirror those of her appointing President, is naturally an oversimplification. However, even if this assumption were generally valid, Professors Choi and Gulati’s measure misses an important dimension of independence. Under their assumption, a judge’s independence from her appointing President could be measured in two ways: the frequency with which a judge dissents from members of her own party and the frequency with which she agrees with members of the opposing party. However, the tournament considers only the former and thereby risks mismeasurement. For example, a highly ideological judge who outflanked members of her own party on the ideological extreme might regularly dissent from decisions joined by members of her own party because she perceived them as too moderate. In so doing, the highly ideological judge would appear highly independent in the tournament. By not considering the regularity with which the judge dissented from members of the opposing party, the tournament may fail to identify a judge who is persistently political.

This hypothetical might be criticized as too far-fetched, but it also points out the conceptual inconsistency in the tournament’s quality and independence measures. For the quality measure to reflect the superiority of legal reasoning and analysis, judges must cite opinions on the basis of their objective worth, independent of their politics. For the tournament’s independence measure to reflect autonomous thinking, judges must be predictably political in deciding cases. The models of judicial behavior underlying the two measures are in tension with one another.

The tournament’s independence measure fails to account for other complexities of judicial behavior. A judge who thinks independently might be more willing to compromise and may dissent relatively infrequently. For that judge, a small change in the number of her dissents would greatly influence how she performed along this measure.

44. See Choi & Gulati, Empirical Ranking, supra note 2, at 61-67.
Similarly, judges who stave off writing dissents by persuading their peers of their own view may be highly independent thinkers, but they would not appear so in this measure. In addition, a judge sitting in a circuit that consistently renders the legally “correct” decision could excel in this measure only by producing dissents of questionable value.

D. Other Dimensions

The tournament is hampered by its exclusion of certain characteristics conventionally thought to be important in judges and Justices. A judge may contribute mightily to the quality of an opinion even if she is not its author. A thoughtful judge may ask penetrating questions from the bench that help shape the views of the other members of the panel. In conference discussions or in commenting on a colleague’s draft opinion, a judge may influence an opinion’s analysis. In circuits with frequent en banc rehearings, a judge may spend significant energy reviewing the court’s decisions and writing to inform her colleagues of the desirability of rehearing particular cases en banc, but the public is never informed of these efforts. These characteristics and efforts are reflected only obliquely in the tournament to the extent that they correlate with the three dimensions measured.

To a good degree, however, all of these criticisms are snipping around the edges of the Choi/Gulati tournament. What is remarkable about their exercise is the dominance of Judges Posner and Easterbrook along the productivity and quality measures, and the bunching together of virtually everyone else. Although the Seventh Circuit duo measures up less well in the independence metric, persons who are at all familiar with their writings—from on or off the bench—would not question the independence of their thinking. Their performance in this category may itself indicate that the quality of independence is extremely difficult to quantify. Their dominance also raises the question of why the tournament is necessary at all. Everyone in the legal community knows that these two judges are the rarest of talents. The question that Choi and Gulati do not ask is why these two do not belong to the so-called “Bush Five,” the rumored short-list of potential nominees. The fact that they are not on the rumored list suggests that the tournament fails to capture some relevant phenomenon.

In their writings off the bench, Judges Posner and Easterbrook have at times been intellectual provocateurs. For example, Judge Easterbrook has criticized mandatory disclosure rules in the federal securities laws\(^45\) and has characterized—not disapprovingly—the

rules of criminal procedure as creating a marketplace. Judge Posner has explored the possible benefits of baby-selling and licenses to commit rape, to name just two examples. These academic writings are noteworthy, even classic, pieces of legal scholarship because they rigorously analyze why the improbable could be possible. A penchant for provocation, sometimes at the expense of an audience’s sensibilities, makes for interesting reading, classroom discussion, and academic symposia, but it might not be a desirable characteristic in a Supreme Court Justice.

The public might prefer its judges and Justices to possess numerous other characteristics that the tournament excludes. For example, the public might want its judges to be fair, genuinely concerned about the parties coming before the court, and humbled by the authority entrusted to them. These qualities—often lumped together under the rubric “judicial temperament”—do not readily lend themselves to quantification. The tournament’s best performers, as well as its worst ones, might have these characteristics in abundance, or they may have them in varying quantities. Because the tournament does not include them, the reader simply does not know.

The bunching together of other judges and the observation that small changes in the number of opinions, citations, or dissents significantly reorder the rankings suggest that the tournament lacks something else. Specifically, it does not contain a measure that permits most judges to be distinguished from one another. This absence may be an inadequacy of the tournament or it may reflect the technology of judging. Perhaps most appellate judges perform at roughly the same level because they face the same technology, possess the same staff resources, and, within a certain range, are armed with the same amount of intellectual firepower. Perhaps the most important similarity in the work of federal judges is that all of them labor with and under the same body of law: the Constitution, federal statutes and regulations, and precedent of the Supreme Court and the respective circuit courts.

The appropriate interpretation of the tournament’s main result—that, with few exceptions, most federal judges perform roughly case for [securities] regulation of any sort, let alone for the particular regulations the SEC uses.

50. See Choi & Gulati, Empirical Ranking, supra note 2, at 75 (“Without Judge Posner or Judge Easterbrook, no one winner emerges.”).
equally—is unclear. Professors Choi and Gulati chose to emphasize a few judges at the top of the ranking rather than the ranking’s inability to distinguish most of the judges. The primary lesson of the tournament might instead be that since one federal judge is about as good as another along the dimensions they measure, the few outliers in the tournament are effectively statistical noise. If so, this interpretation would argue in favor of replacing the current nomination process with a randomization device in which a Supreme Court nominee was chosen by lot from the pool of federal appellate judges. A randomization process would achieve their goal of reducing partisan bickering. Curiously, Professors Choi and Gulati did not consider this alternative.

Ultimately, the exclusion from the tournament of relevant characteristics imparts some measurement error to the prediction of the judge most qualified to sit on the Supreme Court. This inaccuracy implies that a nomination of a judge other than the tournament winner to the Supreme Court may not, as Professors Choi and Gulati contend, be solely a function of ideology.

II. WHAT HAPPENS WHEN JUDGES “JUDGE TO THE TOURNAMENT?”

Professors Choi and Gulati emphasize that the tournament measures a judge’s productivity or work effort. By calling their rankings a tournament, they reference an economics literature in which hierarchical reward structures induce greater effort from workers. The suggestion that appellate judges need competition implies that the current system is not operating optimally. But the idea that appellate judges exert insufficient amounts of effort contrasts with the prevailing wisdom that the federal appellate bench is understaffed. Since 1960, the number of cases filed in federal appellate courts rose by more than fifteenfold, but the number of judges increased by only 2.5 times. As a result, the average active judge handles 956 more cases per year than she did forty years ago. Although caseload var-

51. See Choi & Gulati, Empirical Ranking, supra note 2, at 43 (describing published opinions as requiring more effort than unpublished ones); id. at 62 (emphasizing the effort required to write dissents and concurrences).
52. See Sherwin Rosen, Prizes and Incentives in Elimination Tournaments, 76 AM. ECON. REV. 701, 709 (1986) (explaining that prizes in a sequential elimination tournament must be concentrated at the top in order to induce sufficient effort in later rounds); cf. Edward P. Lazear & Sherwin Rosen, Rank-Order Tournaments as Optimum Labor Contracts, 89 J. POL. ECON. 841, 849 (1981) (suggesting that a tournament induces optimal investment in skills).
54. See id.
ies considerably across circuits, it has risen in all circuits. The crush of cases already provides judges ample incentive to work hard.

Competition among circuit judges is a curious virtue to extol. If the overall effort of appellate judges is not in question, then the need for competition suggests that the distribution of judicial resources has been misallocated and that the tournament would redirect judicial resources in the right direction. It is here that we have our greatest differences with Choi and Gulati’s proposal. It is precisely because the tournament’s qualifying criteria are known to each judge that there is a possibility of competition. We may fairly ask: Do we really wish to foster competition along the productivity, quality, and independence measures? Are there perverse consequences to the competition? What happens when judges “judge to the tournament”?

The structure of the federal appellate courts appears designed to discourage competition among judges. In contrast to the separated branches, which face “hydraulic pressure . . . to exceed the outer limits of [their] power,” within the judicial branch appellate courts are collegial, not competitive, bodies. Circuit courts have a near monopoly over federal cases and controversies within their jurisdiction, limited only by the Supreme Court’s decisions on those questions. Collegial decisionmaking not only encourages but demands collusion (in the economic sense) among judges. The terms of employment for federal judges discourage their thinking about possible promotion to the Supreme Court. Federal appellate judges are not even guaranteed consideration for a nomination because the President may look beyond the federal bench for good candidates. Perhaps most important, Article III of the Constitution grants each judge a lifetime appointment, conditional on good behavior. Life tenure spares a judge the temptation to tailor her decisions to retain her current position or to angle for the next one. Professors Choi and Gulati would undo this system and have judges altering their decisions to jockey for an appointment to the Court.

The first “competitive” consequence of the tournament would be analogous to the “school-qualifying-exam” effect. Exam-based promotion systems in secondary schools have as their goal that students demonstrate minimal competency. Critics contend, however, that the tests create an incentive for educators to maximize passage rates by

56. U.S. CONST. art. III, § 1. Life tenure in a single position contrasts with the usual economic rationale for a tournament. Often an actor’s skills are unknown, and a sequential-elimination tournament provides a cost-effective mechanism for players to determine if they possess the skills necessary to succeed in the endeavor. If they do not, they may quickly exit and seek other pursuits. See Lazear & Rosen, supra note 52, at 861 (suggesting “tryouts” when ability is unknown); Glenn M. MacDonald, The Economics of Rising Stars, 78 AM. ECON. REV. 155 (1988).
ignoring any subject not examined. Similarly, ambitious judges might ignore any activities and abandon any qualities that do not advance their rankings in at least one of the three dimensions of the tournament. For the reasons previously described, this single-mindedness would not serve the public well. Second, even if judges limited their attention to just the criteria of the tournament, they might still modify their decisionmaking in ways that do not further the rule of law or the quality of justice. Perhaps the easiest dimension to manipulate is the productivity ranking. A judge’s relative position improves by designating more decisions for publication. A judge need not actually improve the quality of the decision in order to advance in the tournament’s productivity ranking. Unlike law professors who must persuade journal editors that their writings are worth publishing, Federal Reporter, Third Series, is a circuit judge’s Frostian home: when judges go there, West Publishing has to take them in. A counterargument may be that a judge’s concern with the quality of her opinions, as measured either by the tournament or by her reputation in the legal community, constrains her willingness to publish everything. This argument has some merit, but it falls short in three ways.

First, a recent proposed change in the Federal Rules of Appellate Procedure (F.R.A.P.) has prompted circuits with relatively large unpublished opinion practices to consider how they would respond if every decision were treated as a published decision, that is, as precedent that can be cited to the court.57 The conventional wisdom is that if proposed F.R.A.P. 32.1 were adopted, then a decision, which previously would have contained some analysis and explanation of the decision for the benefit of the parties, will instead issue as a summary affirmance or reversal, albeit a published one. The tournament redoubles the incentive for summary decisions. Because the tournament counts only published opinions without reference to their content, summary decisions boost a judge’s productivity ranking even though they require less effort than the types of unpublished decisions issued currently.

Second, the quality rankings interact with the incentive to publish in odd ways. Citation counts encourage judges to distinguish prior precedents, even to author conflicting opinions, since such opinions will likely be acknowledged by other courts. Take our prior example of the Seventh Circuit’s Booker decision. A court is more likely to cite Booker itself than a subsequent opinion that agrees with Booker and

adopts its approach. However, an opinion that disagrees with Booker or that adds a nuance to its analysis is more worthy of citation. An ambitious judge in another circuit would therefore have an incentive to depart in some way from Booker, even if she agreed with its reasoning and conclusion. By relying on citation counts, the tournament rewards judges for the novelty of their opinions; the result may be less clarity or certainty in the law.

Third, one version of the quality rankings counts just the citations to a judge’s “top twenty” opinions. This scheme strengthens the incentive to issue summary judgments in cases that do not present novel or interesting questions. An ambitious judge could summarily decide the bulk of cases and focus her attention on crafting a few spectacular opinions. In so doing, she would boost her tally of published opinions and jumpstart her quality ranking. The quality rankings create incentives for other strategic behaviors. As described above, the tournament’s quality measure spurs pithy statements and judicial aphorisms. This reward structure is at some tension with the actual quality of legal analysis. In the political sphere, one-liner characterizations of a candidate’s positions are much decried. The tournament rewards the proliferation of this practice in judicial writing, where more thoughtful analysis should prevail. Furthermore, a self-interested judge might collude with other judges to boost her citations. Senior judges who visit other circuits have an opportunity to cite their home-circuit peers in the opinions they author in the visited circuit. The tournament creates an incentive for a judge seeking to boost her quality score to engage in “trades” with a senior judge in exchange for out-of-circuit citations.

Even if a judge cannot increase her own quality rankings, she can suppress those of her rivals in other circuits. The transparency of the criteria imply that a judge knows precisely the identity of the other judges clustered around her and, thus, whose opinions to avoid citing. The tournament creates an incentive for judges to do the opposite of what Professors Choi and Gulati believe makes their quality measure plausible: to cite opinions for strategic reasons rather than for the quality of the legal reasoning and analysis. The tournament itself corrodes the validity of this measure.

The tournament’s independence ranking may have the most predictable incentive effect. To stand out in this measure, a judge should simply dissent more often when she sits with judges of the same political party. Additional dissents and concurrences breed confusion and complexity in the law. The American courts have a tradition of fostering collegial opinions. Among Chief Justice Marshall’s greatest contributions to the Supreme Court was his leadership in effecting

---

58. See Choi & Gulati, Empirical Ranking, supra note 2, at 54-58.
its departure from the English practice of *seriatim* opinion writing.\(^{59}\)

The tournament erodes this tradition. A greater agreement among the panel indicates the judiciary’s solidarity in the basis of its judgment. We recognize this when we debate the viability of 5-4 decisions of the Court, and we acknowledge the unassailability of a unanimous judgment and a single opinion. Perhaps the strongest signal of solidarity that a court can communicate is the unanimous and unsigned per curiam opinion. The tournament weakens the willingness of judges to join a per curiam because doing so forgoes an opportunity to appear independent. It also weakens the incentive to draft a per curiam because it denies the author an additional point in the productivity rankings.

Our criticism of the tournament should not be mistaken for mere cynicism. By suggesting that judges might engage in strategic behavior in response to the tournament, we do not impugn the integrity of the American bench. Our criticism is directed to the criteria for winning the tournament, which, Professors Choi and Gulati tell us, are conspicuously “transparent”\(^{60}\) precisely so that judges will know what they have to do to win, and “gaming is a good thing.”\(^{61}\) What we see as an objection—that judges will compete in the tournament—Professors Choi and Gulati see as the tournament’s virtue. However, as we have described, the tournament creates incentives for a gamesmanship that does not reflect desirable qualities in a judge or Justice, and therefore a whole new set of additional rules might have to be introduced to ensure the tournament actually represents the “fair play” that its creators envisioned.

The establishment of the tournament as a selection device for Supreme Court nominees also changes the incentives facing the President. Professors Choi and Gulati intend the tournament to motivate the President to pick its winner.\(^{62}\) However, the greater the import of the tournament, the greater the temptation to pick a candidate from outside the federal appellate bench. Sidestepping the tournament altogether is the same maneuver that some Presidents have purportedly taken to eschew scrutiny of a nominee’s views on hot-button issues. As previously described, a judicial opinion on a hot-button issue may provide insights into a nominee’s ability. A candidate who has not written on these issues is more difficult to evaluate, and hence to oppose, because relatively less is known about her. Similarly, a candidate who is unranked in the tournament cannot be compared to the


\(^{60}\) Choi & Gulati, *Empirical Ranking*, supra note 2, at 36.

\(^{61}\) *Id.* at 34.

\(^{62}\) *See id.* at 27-28.
set of alternative candidates. By picking a nominee who is not ranked in the tournament, a President could avoid precisely the scrutiny that the tournament is intended to apply. Rather than expanding the amount of information available about nominees, the tournament encourages the selection of “stealth” candidates or candidates without paper trails.

The tournament also alters the public’s incentives. The ranking of judges creates a new set of data easily grasped by the public. A President who nominates number forty-two on the list will have to explain himself. And even if the top ten on the list are not feasible candidates for one reason or another, the President will still have to explain why judges ranked eleven through forty-one are not available for nomination. In contrast, he will not have to explain why he did not nominate the judge ranked forty-third because the reason will be obvious to everyone. That is a heavy responsibility to place on a ranking. Additionally, the content-free nature of the tournament makes it easier for the public to have an “informed” opinion about a nominee’s fitness for the Court without acquiring information about the nominee’s judicial philosophy. Like baseball fans who read box scores and participate in rotisserie leagues but never watch any games, students of the tournament may vigorously debate a nominee’s worth without reading a single opinion. The tournament would make it less, not more, likely that discussions of the desirability of a particular nominee would be based on a review of the judge’s actual work product.

III. SHOULDN’T WE COMPARE APPLES TO APPLES AND ORANGES TO ORANGES?

As other contributors to this Symposium have observed, the qualities that allow a judge to excel on a circuit court are not necessarily those that engender success on the Supreme Court. It is not self-evident why productivity on the courts of appeals, citation count, and independence are the qualities we most need in a future Supreme Court Justice. Just to take one example, productivity may be a proxy for diligence, and we can agree that we generally want diligent Justices, but the Supreme Court’s docket has been shrinking over the past years from an average of about 170 cases to roughly ninety-five cases per year. Is the productivity measure an effort to boost flag-

---

63. See id. at 28.
64. Mark Tushnet, Foreword: The New Constitutional Order and the Chastening of Constitutional Aspiration, 113 Harv. L. Rev. 29, 68 (1999) (“At its peak, the Court heard argument in about 170 cases per Term, whereas during the past five years it has heard argument in fewer than one hundred cases each year.”); see also Hart v. Massanari, 266 F.3d 1155, 1177 n.34, 1178 n.37 (9th Cir. 2001) (noting that in the 1999 Term, the United States Supreme Court heard seventy-seven cases, or less than nine majority opinions per Justice,
ging productivity of the Court? Or should the public regard a Supreme Court appointment as a well-earned retirement after hard work on the circuit?

Professors Choi and Gulati have exhaustively studied which apple is best on the assumption that it will make the best orange. The winner of the tournament has himself acknowledged the possibility that greatness as a circuit judge does not necessarily assure greatness as a Justice.65 This observation is not just a criticism of the tournament but instead a suggestion for another use to which it might be put. Rather than using the measure to predict which appellate court judge would make the best Supreme Court Justice, the Choi/Gulati metric might be useful in identifying whom should be nominated to the circuit courts in the future.

If a consensus existed that the tournament accurately measured the desired characteristics of circuit judges—a strong assumption—then it might be used to identify the traits that impart success to circuit judges. Why are the judges that succeed in the tournament so much better than the rest of the pack? Academics have studied and debated the relationship between a judge’s personal characteristics and the outcomes of particular types of cases.66 A similar exercise might be undertaken with the tournament. A judge’s score in the tournament might be related to her individual characteristics. Such a study might reveal that the best judges are lapsed academics, erstwhile appellate practitioners, or former district judges. Or, perhaps only a former editor-in-chief of her law school’s law review thrives on the bench. If so, such information would help the President and Senate identify and evaluate candidates for the appellate bench. Whether such a study would reveal a short set of qualities that assure success on the appellate bench is uncertain. Moreover, it is not clear that we want a bench comprised of judges satisfying a narrow judicial “profile.” We have more democratic tradition in this regard than other countries whose judges rise through the civil service and undergo rigorous, uniform training. Democracy can be unruly, and as evidenced by our history, we have long assumed that the judiciary’s membership should reflect a variety of life experiences. However,


such a study would at least compare apples to apples, rather than apples to oranges.

If interest instead lies in what makes a successful Supreme Court Justice, the appropriate subjects of study are the Justices themselves. The task of measuring performance of Justices on the Supreme Court is no small feat, as other contributors to this Symposium have discussed.\textsuperscript{67} If a measure of a Justice's success on the Court existed, however, it could be related to his or her characteristics in a fashion similar to that described for circuit judges above. The results of the studies of circuit judges and Supreme Court Justices could be compared to determine whether the individual characteristics that bred success in one position did so in the other. Although this proposal is a wholly different research enterprise than Professors Choi and Gulati's, it would test directly their strong assumption that traits that make for a good circuit judge also make for a good Justice. It would force us to rethink the criteria: Do we believe that productivity, citation counts, and independence are the hallmarks of the great Justices? How do the great Justices fare in a "Tournament of Justices?"\textsuperscript{68}

The failure to test the key assumption of their model illustrates that the tournament, although it involves measurement, is not really social science. Ordinal rankings, like David Letterman's Top Ten lists, or numerical measurements, like baseball batting averages, are mere quantification. For the reasons described previously in this Essay, whether the tournament—despite its elaborateness—actually captures the phenomenon it purports to is uncertain. The use of a measurement whose validity is untested imposes three costs. First, as previously discussed, the measure may distort the incentives of the subjects who know that they are evaluated according to this metric. Second, judges who rank low in the tournament because of its imprecision are unfairly besmirched. Professors Choi and Gulati pull their punches by claiming that they cannot infer anything about the low-ranking judges.\textsuperscript{69} If so, then perhaps only top finishers in the tournament should be announced. Third, the tournament also harms the public. Should litigants who appear before lower-ranking judges


\textsuperscript{69.} \textit{See} Choi & Gulati, \textit{Empirical Ranking}, supra note 2, at 75-77.
conclude that they have received an inferior brand of justice? If Professors Choi and Gulati believe their tournament accurately measures the adequacy of judicial performance, are they suggesting that litigants appearing before lower-ranking judges did not receive due process?

IV. CONCLUSION

In complex human endeavors, attempts to quantify the dimensions of success are inherently appealing. Sports fans who keep game statistics are a prime example. But what makes rankings so enjoyable to discuss and analyze is that at some level fans know that the numbers never measure all the aspects of an athlete’s play. Some intangible residual is left out, and this left-out portion might be the most important feature of an athlete’s play. In judging, where outcomes are even less readily observed than sports, the gap between numerical measures and actual performance is greater. Professors Choi and Gulati’s attempt at measurement is a valiant one, but to some extent, they have told us what we already knew: Judges Posner and Easterbrook are exceptional talents, and the rest of the appellate bench is hard to distinguish from one another. Their measures are imprecise ones, and were judges to compete along these dimensions, society might bear considerable costs. Studies that expand our knowledge of the work of the judiciary deserve applause, but Professors Choi and Gulati have not yet presented a convincing case that choosing Supreme Court nominees by tournament should replace selecting them by the Constitution’s more political process.
WHICH JUDGES WRITE THEIR OPINIONS (AND SHOULD WE CARE)?

STEPHEN J. CHOI* AND G. MITU GULATI**

PROLOGUE: LAWYERS AND THE ATTRIBUTION OF AUTHORSHIP

If one of our students were to pass off someone else’s work as his or her own and we were to discover that misrepresentation, the penalties would be harsh. But as lawyers, we hold ourselves to less rigorous standards for plagiarism than those to which we hold our students. Where research assistants make significant contributions to

* Professor of Law, New York University School of Law.
** Visiting Professor of Law, University of Virginia School of Law; Professor of Law, Georgetown University Law Center. We are indebted to David Keeney, Robert Mezey, and Craig Hoffman for conversations about linguistic theory, authorship styles, and how one might track and measure style differentials in judicial opinions as opposed to fiction or poetry. We are grateful to the computer scientists and programmers at InApp in India, including Satish Babu, M.C. Jayakrishnan, and R.V. Suchitra, for their innovations in designing programs to run standard authorship tests on judicial opinions. For conversations about the ideas here, we also thank Gail Agrawal, Scott Baker, Michael BeVier, Devon Carbado, Adrienne Davis, Susan French, Dennis Hutchinson, Edeanna Johnson, Kimberly Krawiec, Richard Posner, Un Kyung Park, Tom Rowe, Michael Solimine, David Vladeck, and P. Vijaykumar. Our thanks also go to Associate Dean Jim Rossi and the Law Review editors at the Florida State University College of Law for organizing this Symposium. The library staffs at the University of Virginia School of Law and the Georgetown Law Center provided us with invaluable research assistance. Chris Knott and Kent Olson were especially generous with their assistance. Finally, we owe a debt to Lee Epstein for giving us the initial impetus for the project.
their professors’ papers and sometimes even draft large sections of those papers, they are rarely given anything more than an acknowledgement in a footnote.¹ Law firm partners, we suspect, think nothing of asking junior associates to draft entire articles or book chapters and then send them out under their name (again, often without acknowledgement of the flunky’s authorship). And then there is the matter of judges. Law clerks are said to draft the vast majority of opinions for judges. Yet, if one were to ask most lawyers and judges whether authorship credit should be given to the individual clerks, they would in all likelihood think the question ridiculous. Why do we, as lawyers, consider proper attribution of authorship so important for our students and so unimportant for ourselves?²

I. INTRODUCTION

Federal judges enjoy a large amount of discretion in how they go about writing opinions. Once a judge is assigned an opinion, the judge may choose to write the opinion alone, doing both the research and writing without any assistance. Judges may also turn to their clerks to help research relevant law or to draft parts of the opinion. Some judges may allow clerks to draft entire opinions, while they themselves contribute only a light editorial overview. Commentators have discussed the practice of delegating significant portions of the opinion-writing task to clerks, and more than a few have criticized it.³ But no one seems to know exactly how much delegation goes on;

¹. Law professors, their misuse of sources, and failures to properly attribute credit have been in the news lately. See Daniel J. Hemel & Lauren A.E. Schuker, Prof Admits to Misusing Source, HARV. CRIMSON ONLINE, Sept. 27, 2004, at http://www.thecrimson.com/article.aspx?ref=503493 (quoting Harvard Law Professor Alan Dershowitz as making the justificatory point that although the rules governing plagiarism by undergraduates are clear (and harsh), the norms of attribution and citation in the legal profession—where judges frequently rely on lawyers’ briefs and clerks’ memoranda in drafting opinions—are less clear); Sara Rimer, When Plagiarism’s Shadow Falls on Admired Scholars, N.Y. TIMES, Nov. 24, 2004, at B9 (discussing criticisms of Harvard Law Professors Laurence Tribe and Charles Ogletree Jr.).

². See Marilyn V. Yarbrough, Do as I Say, Not as I Do: Mixed Messages for Law Students, 100 DICK. L. REV. 677 (1996) (discussing the mixed messages about plagiarism that law professors and legal professionals send to law students).

commentators rely on anecdotes, rumors, and, at best, informal surveys.4

In prior work, we asked why, in selecting Supreme Court Justices, there was little attempt to use the available data on the relative performances of federal circuit court judges (the primary pool from which Supreme Court Justices are chosen).5 In this Essay, we ask why few have attempted to provide a systematic analysis of authorship patterns among federal circuit court judges. We use generic techniques from computational linguistics, as well as several methods tailored for the judicial setting, to explore both the desirability and feasibility of determining the authorship of judicial opinions.

Information about the production process (as opposed to information about the end product alone) can be useful in decisions regarding judicial promotion. Whether a particular judge expends effort in authoring her own opinions may have bearing on how suited the judge is for elevation to a higher court, including the Supreme Court. Knowing whether individual judges use their scarce time to manage their clerks’ writing or to engage in the writing task themselves may help determine whether there is a need for more judges and resources for the judiciary. If we observe that a spike in the volume of cases coincides with judges suddenly placing greater reliance on their clerks to write opinions, we might want to support an expansion of the federal judiciary. Authorship information may also help identify judges who are no longer able or willing to perform their tasks adequately. Once identified, peer pressure and other forms of public opprobrium may lead the judges to either increase their activity level or

4. As the reader will later see, the criticism about the reliance on anecdote, rumor, and informal surveys is something that our project is also subject to since we needed some baseline piece of information about “true” authorship levels to test out various authorship testing methods. The goal, however, is to develop a set of methods of testing authorship such that reliance on these often nonverifiable methods can be reduced.

retire. Law students applying for judicial clerkships will find it useful to know whether their potential employer does her own writing or delegates it all to her clerks. And—highly relevant from our perspective as researchers—authorship information has the potential to improve academic research on judicial behavior.

Part II of this Essay sets out a framework to explore whether it is worthwhile to investigate which judges write their opinions. We explore whether there is societal value in knowing information about the input levels that society’s agents (in this case, the judges) put into the production process (the product being judicial opinions).

A variety of techniques may be employed to determine the authorship of opinions. Part III sets out tests to determine the viability of these techniques. We use the limited existing information on authorship—a handful of judges such as Richard Posner, Frank Easterbrook, and Michael Boudin participate more actively in the writing of their opinions than do most other judges6—to test how well the different authorship tests perform in distinguishing such judges.

Our sample pool consists of all circuit court judges who were both active and under the age of sixty-five as of May 2003 and had been on the bench for the period from January 1, 1998 to December 31, 2000.7 For each judge in the sample, we selected opinions for analysis at random from those generated during the three-year sample period.

Based on this sample, we report that the generic tests drawn from computational linguistics fail to distinguish Posner, Easterbrook, and Boudin as judges most likely to author their opinions. The generic tests, however, do not control for the subject matter of specific opinions. The common phrases used in opinions of a specific genre (for example, administrative law opinions) will cause the generic methodologies to treat all such opinions as more likely to be by the same author, even if different authors actually wrote the opinions. If the randomly selected opinions for one judge are all of the same subject matter but those for another judge are not, this factor alone may lead

6. Several commentators have remarked on the writing practices of Judges Posner and Easterbrook. See Robert F. Blomquist, Dissent, Posner-Style: Judge Richard A. Posner’s First Decade of Dissenting Opinions, 1981-1991—Toward an Aesthetics of Judicial Dissenting Style, 69 Mo. L. Rev. 73, 74 n.12 (2004); Martha Middleton, Shaping a Circuit in the Chicago School Image, Nat’l L.J., July 20, 1987, at 1 (discussing Easterbrook and Posner). Informal conversations with a number of other judges and former law clerks confirmed that Judges Posner and Easterbrook author all of their own opinions. Many also mentioned Judge Boudin as someone who authored most of his own opinions. Other judges were also mentioned, albeit less frequently, as authoring substantial portions of their opinions. We use Posner, Easterbrook, and Boudin—those judges with the highest a priori likelihood of self-authorship—to calibrate the effectiveness of our authorship methodologies.

7. We generated much of this dataset in our earlier article ranking judges based on judicial performance. See Choi & Gulati, supra note 5. To take advantage of some of the data collected for that project, we restrict our analysis to the same pool of judges here.
the generic authorship tests to give the first judge a higher self-authorship score.

We also provide more customized tests designed to control, at least in part, for the subject matter of judicial opinions. Using these tests, we are able to distinguish our test judges—Posner, Easterbrook, and Boudin—as ranking consistently high in terms of self-authorship of judicial opinions. Part IV concludes and provides possible extensions of the authorship methodology.

II. THE UPSIDES AND DOWNSIDES TO DETERMINING JUDICIAL AUTHORSHIP

The production of judicial opinions is a joint venture between the judges and their staffs—for purposes of opinion writing, the law clerks. The outside world sees the end product in the final opinion that appears in West’s case reporters. Outsiders cannot distinguish the contributions of the various participants. For each opinion, only the primary writing judge is identified (for an appellate opinion, the other judges on the panel are also identified as secondary participants, but their relative involvement is not specified).

Lore has it that many opinions are drafted primarily by the judge’s law clerks and sometimes even by staff attorneys. Such delegation may help judges handle their ever-increasing caseloads. At the other extreme, there are some judges who, despite the caseloads, are reputed to take complete responsibility for opinion writing (for example, Richard Posner and Frank Easterbrook, who seem to be able to do this and more). Further, occasions may arise when the other judges on the panel draft significant portions of the opinion. But the involvement of these subsidiary coauthors is never identified. In addition, norms of appropriate behavior mean that the clerks rarely reveal what occurred in their judges’ chambers; the information the clerks possess about relative responsibility for the various opinions is seldom reported in any public and verifiable manner.


10. When clerks do reveal the goings-on in their judges’ chambers, they can sometimes receive considerable criticism. A recent example is the criticism that was leveled at Edward Lazarus’ Closed Chambers, a memoir of his year on the Supreme Court clerking for Justice Blackmun. See Kenney, supra note 3 (discussing some of the criticism). Prior to that, there was criticism of the many law clerks who spoke to Bob Woodward and Scott Armstrong when they were researching their book on the workings of the Court, The Brethren. See George Gold, Loose Tongues: How Stone Cautioned Clerks, A.B.A. J., Oct. 1985, at 28; Mahoney, supra note 3, at 336.
The agency concept provides a useful framework for thinking about the judicial production problem. Judges are agents who perform a set of tasks, which include producing judicial opinions, for the public (the principal in the relationship). A variety of consumers—lawyers, law students, litigants, researchers, and other judges, among others—use these judicial opinions to understand the law. In order to perform their tasks, judges are authorized to employ a set of subsidiary agents (law clerks and staff attorneys), who are hired, supervised, and evaluated by the judges themselves. Society's goal as the principal is arguably to ensure that the judges produce at maximal potential, while at the same time providing for the judges' independence.

Ordinarily, principals attempt to monitor the conduct of their agents. If the agents do not work hard enough or produce high-quality products, they get fired (or, in the case of politicians, they get voted out of office). Since society wishes to ensure the independence of federal judges, however, they are provided with virtually ironclad job security and fixed salaries. These elements of job security and fixed incomes mean that the public has relatively few levers with which to incentivize the judges; for the most part, judges do their work because they want to. It may appear pointless to inquire into judges' relative levels of involvement in the production of opinions. Indeed, that view may be one reason why such information thus far has not been collected.

As illustrated by the materials mentioned above, the one context where there has been some minimal revelation of the levels of delegation to law clerks is the Supreme Court. Even here the information is highly imperfect, often a function of clerk reconstructions in contexts where the explicit goal is to praise the Justice or journalistic reports about high profile cases resting on undisclosed sources. But at least there is some information. Plus, to the extent that some former Justices have (a) opened their papers to the public and (b) kept accurate records on clerk-judge communication, researchers have a starting point to try and determine the level of clerk delegation. Among the prominent recent examples along these lines are the Legal Affairs article by David Garrow criticizing the amount of delegation of power by Justice Blackmun to his clerks (especially later in his career) and the Vanity Fair exposé, following the 2000 election, that used information from law clerks to construct a story about how the decision was made. See David Garrow, The Brains Behind Blackmun, LEGAL AFF., May-June 2005, available at http://www.legalaffairs.org/issues/May-June-2005/feature_garrow_mayjun05.msp; David Margolick et al., The Path to Florida, VANITY FAIR, Oct. 2004, at 310. For examples of former law clerks writing about their Justices, see JOHN C. JEFFRIES, JR., JUSTICE LEWIS F. POWELL, JR. (1994); J. HARVIE WILKINSON, III, SERVING JUSTICE: A SUPREME COURT CLERK'S VIEW (1974); Bruce A. Ackerman, In Memoriam: Henry J. Friendly, 99 HARV. L. REV. 1709 (1986); Anne M. Coughlin, Writing for Justice Powell, 99 COLUM. L. REV. 541 (1999); James L. Volling, Warren E. Burger: An Independent Pragmatist Remembered, 22 WM. MITCHELL L. REV. 39 (1996); and Kevin J. Worthen, Shirt-Tales: Clerking for Byron White, 1994 BYU L. REV. 349. The precise contours of what can and cannot be revealed by former clerks, however, seems to be unclear. That lack of clarity of the disclosure rules, when combined with the fondness (or fear) that most clerks have for their judges, we suspect, keeps most clerks from going on the record with information about the authorship practices in their judges' chambers.
In Part II.A, we put forward the claim that a better understanding of the level of judicial input into the opinion-writing process can potentially help the management of judicial agents in at least three circumstances: deciding on promotion when the quality of the final output is hard to evaluate, determining incentives for the judges as part of a judicial opinion production team, and assessing how best to allocate resources to the judiciary. We also identify others who may find knowledge on authorship useful: researchers studying judicial behavior, law students entering the judicial clerkship market, and those considering how much to rely on the opinions of specific judges.

Having considered the upsides, in Part II.B we note possible downsides to the project that colleagues have mentioned to us. These are the danger of improper negative inferences and the problem of imperfect measurement.

A. Upsides

1. Promotions and the Quality of Judicial Output

Information on input into the production process can assist in the evaluation of the quality of a final product where such quality is difficult to observe directly. For many products, such as legal and medical services, product quality is hard to measure. When a lawyer loses a case or a transaction fails to be completed, it is difficult to know whether the result should be attributed to the lawyer or to other factors. If a patient’s health does not improve under a doctor’s care, does this reflect the doctor’s failure to deliver competent medical care or the patient’s own initial health condition?

Where the final output is difficult to evaluate, another solution is to look to inputs.11 For example, the difficulty in evaluating lawyers’ product is often given as a reason for why lawyers bill by the hour.12 If one can know the levels of effort and skill a lawyer brings to the production process, one can determine whether the bad outcome was the product of inadequate effort and skill, some other factor, or random chance. While doctors are typically not paid on an hourly basis, information on how much attention and effort a doctor vests in a particular patient is likely to be relevant to any potential medical malpractice claim against that doctor.

In earlier work, we suggested that one way to evaluate the quality of a judicial opinion was to look at citation rates.\footnote{See Choi & Gulati, supra note 5, at 34, 48-61.} Citation rates, however, are an imperfect measure because they are affected by a number of factors other than quality: the subject matter of the case, the reputation of the circuit, and the reputation of the judge.\footnote{There is an extensive literature on the pitfalls and benefits of using citation rates as a measure of quality. For discussions, see Arthur Austin, The Reliability of Citation Counts in Judgments on Promotion, Tenure, and Status, 35 Ariz. L. Rev. 829 (1993); Choi & Gulati, supra note 5, at 48 n.38, 54-55; William M. Landes et al., Judicial Influence: A Citation Analysis of Federal Courts of Appeals Judges, 27 J. Legal Stud. 271 (1998); and Russell Smyth, Do Judges Behave as Homo Economicus, and if So, Can We Measure Their Performance? An Antipodean Perspective on a Tournament of Judges, 32 Fla. St. U. L. Rev. 1299 (2005).} Therefore, knowing the level of the judge’s participation in the production process supplements information about citation rates. A judge who rates a low self-authorship score coupled with a low citation rate may signal her relatively low level of engagement in the judicial process.

Information as to individual inputs into the production process can also be useful in instances where the employer cares about more than product quality. An employer might want to know, for example, whether a particular employee is exerting herself fully or only partially. This is information from which the employer can potentially infer a number of things, such as the employee’s commitment to the job, her interest in the project, and her likely effect on her coworkers (to the extent they take their cues from her).

For some jobs, process—which might be seen to include the agent’s commitment and dedication to the process—might have an importance of its own. In the judicial context, commitment and dedication to the task may suggest a concern for justice. Someone who will work day and night on her task, even if she is not the quickest at it, may be more likely to produce a just decision than someone who is highly intelligent and writes quickly but is not willing to spend more than a few hours a day on a case. The willingness to work hard, we suggest, may correlate with a judge’s concern that a case be determined in a fair manner (as opposed to a slovenly willingness to decide the case more randomly). Achieving justice, vague and amorphous a concept as it may be, may be more important to society than ensuring that judges write high-quality opinions.\footnote{See John V. Orth, Judging the Tournament, Jurist (Apr. 15, 2004), at http://jurist.law.pitt.edu/forum/symposium-je/choi-gulati-orth-taha.php.} The same kind of argument can be made with respect to other abstract concepts such as honesty and integrity.\footnote{For example, one could question the honesty and integrity of someone who chose to exert little effort in making a decision even though the lack of effort made it more likely that the outcome was more random and only loosely related to the facts of the specific case.}
Consider the problem of deciding judicial promotions. If the job at the lower level is substantially different from that at the promoted level, high-quality production at the lower-level job may not predict high-quality production after promotion. The work of trial judges, for example, is primarily in the area of trial management, with the occasional authorship of an opinion. Circuit court judges, on the other hand, focus to a greater extent on the production of judicial opinions. Where a trial judge is being considered for promotion to a circuit court position, information about dedication to the job (the inputs) might be more important than which employee produced higher-quality opinions at the lower level (the outputs).

Almost all the candidates given serious consideration for appointment to the Supreme Court tend to have prior judicial experience, most of them on the federal circuit courts of appeals. What, then, if we were to find out that a judge who was a candidate for promotion to the Supreme Court wrote none of her own lower court opinions, but instead delegated them all to her law clerks? And what if there was another candidate who wrote all of her own opinions? This information may not be dispositive in terms of eliminating the first candidate, but questions would be raised and inquiries would be made. If it turned out that the first judge spent a significant portion of the year skiing, sailing, or rock climbing, the public would probably conclude that this judge was not committed to the job.

On the other hand, it might turn out that this judge, instead of actually writing her own opinions, was spending her time managing her clerks and training them to write and analyze. If it also turned out that these clerk-written opinions were considered to be of high quality by experts, the final evaluation of this candidate might be positive. This would be especially so if the job at the higher level required even more management and training skills than the lower-level job. Indeed, the inability of the other judge to effectively utilize her clerks in the writing process might be seen by some as a negative factor in the promotion decision.

The information about the judge’s involvement in the writing process alone is therefore unlikely to be dispositive. It may, however, lead people to look for other pieces of information so as to be able to make inferences. Judges themselves may voluntarily reveal more information about their style of decisionmaking in response to objective information on their involvement in writing opinions. Some evaluators may see direct judicial involvement in the writing of drafts as of

---

paramount importance. Others may regard judges as also having a responsibility to train law clerks. The information that may emerge on judicial input into the writing may be important to both sets of evaluators, although it may lead them to different conclusions about who should be promoted.

2. Fine-Tuning Incentives

A much-discussed question in economics is how to devise solutions to the “team production” problem.18 The team production problem occurs when production occurs in teams and the relative levels of credit for the final product are hard to allocate.19 Where individual responsibility cannot be allocated, a free-rider problem may arise whereby team members have an incentive to shirk.20 There are methods employers use to solve the team production problem, which include internal peer pressure within the team and incentives to members to monitor and manage the other team members.21

In the case of judicial opinion writing, the judge is the manager. The judge, who is allocated full credit for the output of her chambers, has both the ability to monitor subordinates and an incentive to ensure that they work hard. The potential problem, however, is with the judge. If a judge wishes to shirk, there is no formal mechanism to penalize her. A judge can monitor clerks and fire them or give them bad references. But there are few direct mechanisms that enable the public to monitor and penalize a federal judge.

Informal mechanisms, however, do exist. Judges care about status and prestige.22 They will care if information is released showing that some of them demonstrate no involvement in the opinion-writing process (especially if the data also shows that their colleagues are highly involved in that process). Information concerning a lack of authorship, when coupled with a low citation rate, may indicate that a judge is simply not doing a good job, thereby reducing that judge’s reputation among her peers and the general public. Judges may demonstrate greater participation in the authorship process to prevent or remedy this situation.

But what if these judges are not good writers? Might not it be better that they remove themselves from the writing portion of the pro-

18. The classic papers on the team production problem are Armen A. Alchian & Harold Demsetz, Production, Information Costs, and Economic Organization, 62 AM. ECON. REV. 777 (1972), and Bengt Holmstrom, Moral Hazard in Teams, 13 BELL J. ECON. 324 (1982).
19. Alchian & Demsetz, supra note 18, at 779.
20. Id. at 780.
22. Smyth, supra note 14, at 1306-07 nn.39-42 (citing support on the point).
ject? The point is a fair one, but it leads to a more important point: the motivation of politicians. Once a judge who has no skill at writing is appointed, there may be a valid argument that the judge should not be compelled to participate in the writing process. But writing skills and the willingness to write opinions should be at least two of the criteria for appointing a federal judge. And if someone who lacks those basic characteristics is appointed, the blame should fall on the politicians who made the appointment. If politicians are penalized for bad appointments on basic matters such as writing skills, they may be discouraged from using judicial appointments as political favors or from allowing their choice to be determined on the basis of litmus tests; for example, a judge’s likely vote on issues such as abortion and the death penalty. Politicians may then in fact focus on more mundane, yet fundamental, matters such as writing skill.

Another indirect incentive exists. Like the rest of us, judges become ill, old, and sometimes uninterested in their jobs. Being a judge, however, brings status, power, and guaranteed income. There is the danger that a judge who is beyond the point of being able or willing to perform the job will be tempted to eschew retirement. So long as the judge has able law clerks, casual observers on the outside will find it difficult to detect the drop in performance. It is likely that the law clerks and the other judges will be able to observe when the judge is not able to perform adequately, but while there may be internal grumblings in the courthouse, this information will rarely find its way to the general public. We assume that there is a strong social norm against the disclosure of such information.

If, however, information is available as to judicial input into the writing process, any significant drop in interest or capability will result in public pressure for the judge to retire. For example, it is likely that no litigant wants a case to be decided by a senile or otherwise absent judge. And while the law clerks making decisions for these

23. Alternatively, some might argue that judges who are not good writers are precisely the ones who should be incentivized to write more so that they learn to become better writers. In his diary for the online magazine, Slate, Judge Posner writes:

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


24. Cf. Smyth, supra note 14, at 1301-03 (pointing to evidence suggesting that judicial retirement rates respond to changes in incentives, such as pension levels).


26. Supreme Court Justices are perhaps an exception to the rule, as we see from the rumors about Justice Thurgood Marshall. See infra note 42 and accompanying text. Our primary interest in this Essay, however, is in the less visible lower court judges.
absent judges may have done well on their law school exams, the litigants might prefer to have their cases decided by judges who have at least some meaningful experience in the world rather than by students fresh out of law school.

The question that remains is as follows: Would disclosure of this information have any effect on a particular judge, a superannuated one for example? We suspect it would. While there are few external sources of pressure that can be brought to bear on a judge, status, power, and prestige are all a function of the public’s perception of that person. If the public begins to lose confidence in the judge because of evidence that her ability to write has declined, the judge may consider either increasing her effort or retiring.

3. Allocation of Resources

The determination of resource allocation for the federal judiciary is made by the legislature. Given that there are no meaningful competitive forces to tell the legislature when the production process is faulty, the legislature itself has to collect the necessary information to make such evaluations. For example, it is well documented that there has been a dramatic explosion in the caseloads carried by federal judges over the past few decades.27 The number of judges, however, has not kept pace with the increased workload.28 This disparity has led some to call for more judgeships;29 these calls have, in turn, been countered by others.30

One element in the argument that more judges are needed has been that judges today, unlike judges decades ago, are forced to turn to law clerks, staff attorneys, and other shortcuts to enable them to tackle their expanded caseloads.31 Skeptics, however, say that the caseloads are not so large as to prevent existing judges from tackling them. Some of the questions in the debate might be answered by data

27. See COMM’N ON STRUCTURAL ALTERNATIVES FOR THE FED. COURTS OF APPEALS, FINAL REPORT 13-17 (1998); see also Jeffrey O. Cooper & Douglas A. Berman, Passive Virtues and Casual Vices in the Federal Courts of Appeals, 66 BROOK. L. REV. 685 (2000) (discussing the caseload explosion and the responses of the courts of appeals); Lauren K. Robel, Caseload and Judging: Judicial Adaptations to Caseload, 1990 BYU L. REV. 3 (describing and critiquing the methods judges have used to deal with increased caseloads).


31. See sources cited supra note 8.
that would indicate whether judicial involvement in the writing process has decreased in direct correlation with the increase in caseloads over the past several decades.

Court administration can also be helped in other ways. Along with the explosion in caseload, there has been an increase in the complexity and volume of many areas of federal regulation.\textsuperscript{32} In order to tackle this increased complexity and volume, judges may be focusing more attention on certain types of cases while delegating others. If certain types of cases—for example, immigration, social security, securities regulation, or habeas cases—are consistently delegated to the clerks or the staff attorneys (as measured through a low judge self-authorship score), such evidence supports the argument that these specific types of cases deserve a separate set of specialized courts with expert judges, much like the U.S. Tax Court.\textsuperscript{33}

4. Research on Judicial Behavior

Perhaps the most immediate application of information on authorship rates is improvement of the quality of research on judicial behavior. Authorship rates can be used as an explanatory variable in regressions that seek to explain a large number of outcome variables.


\textsuperscript{33} We are not aware of any systematic research into the question of whether there is a greater use of clerks in some substantive areas of the law, although it strikes us that this is an important area of research. \textit{Cf. Ward & Weiden, supra} note 3 (manuscript at ch. 5, at 38-39) (finding some evidence that Justices were more likely to make substantial revisions to draft opinions when particular issues were involved). There are statements by commentators that suggest that the level of clerk and staff attorney usage is much higher in unpublished opinions. See Patricia M. Wald, \textit{The Rhetoric of Results and the Results of Rhetoric: Judicial Writings}, 62 U. CHI. L. REV. 1371, 1373 (1995) (noting that clerks—not judges—generally draft the opinions); Tony Mauro, \textit{Difference of Opinion}, LEGAL TIMES, Apr. 12, 2004, at 1, 10 (quoting Judge Kozinski on the matter). If it turns out that certain categories of cases are more likely to be the subject of unpublished opinions, it should follow that clerk and staff attorney usage will be higher in this category of cases. So, for example, if ninety percent of prisoner or immigration appeals result in unpublished opinions, it may well be that the major portion of lawmaking in these areas is done by law clerks and not judges. This is an outcome that strikes us as both unsatisfactory and calling out for reforms such as specialization. For a trenchant critique of the current practices of issuing unpublished opinions, see Penelope Pether, \textit{Inequitable Injunctions: The Scandal of Private Judging in the U.S. Courts}, 56 STAN. L. REV. 1435 (2004). On the matter of whether the federal courts should specialize more in terms of subject matter, see Jeffrey W. Stempel, \textit{Two Cheers for Specialization}, 61 BROOK. L. REV. 67 (1995).

Along the lines mentioned above, an empirical study by one of the Symposium participants, James Brudney, finds, among other things, that Supreme Court Justices seem to apply statutory canons of construction very differently in cases that arise in the more complex and less high-profile areas of labor law than they do in what might be seen as the more \textit{sexy} areas. See James J. Brudney & Corey Ditslear, \textit{Canons of Construction and the Elusive Quest for Neutral Reasoning}, 58 VAND. L. REV. 1, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=534982.
Among these outcome variables that researchers might be interested in better explaining are voting patterns, citation rates and styles, invocation rates, publication patterns, independence levels, and choices about styles of argument (for example, whether one prefers the use of multifactor balancing tests). It may be that whether a judge writes a significant portion of her opinions is a positive or negative explanatory factor for one or more of these outcome variables. And that information will advance our knowledge of judicial behavior. For example, it has been suggested that there are certain types of cases (perhaps those raising important constitutional questions) where judges are more likely to be actively involved in the writing and editing of an opinion than other types of cases such as tax, securities, and bankruptcy. If authorship rates for different opinions were known, this suggestion could be tested. And, as noted earlier, if it turned out that certain types of cases were consistently delegated to the clerks, this might be reason enough to think about specialized courts for those cases. More broadly, if higher levels of authorship predict higher citation rates (and if citation rates are considered a good proxy for opinion quality), that might suggest that judges should be given incentives to write more of their own opinions. If, instead, regressions reveal the reverse effect, judges might be encouraged to write fewer opinions. Alternatively, if citation rates were seen as an inadequate measure of judicial output quality, one could use whatever other measure (perhaps a survey of experts) and then observe whether higher-quality argumentation or explanation in judicial opinions is related to self-authorship rates. The point is that we will not really appreciate the value of knowing authorship rates

---

34. Invocations involve situations where one judge invokes the name of another judge when discussing an opinion written by that other judge. Choi & Gulati, supra note 5, at 58. We consider invocations as a special sign of respect on the part of one judge for another. See id. at 58-59 (discussing the concept of invocations).

35. One way of measuring judicial independence is to examine whether a judge tends to side more systematically with judges nominated by a President of the same political party as the judge in question. See id. at 63. We analyze circuit court judges, active from 1998 to 2000, based on this measure of independence in Choi & Gulati, supra note 5, at 61-67.

36. See Ward & Weiden, supra note 3 (manuscript at ch. 5, at 38) (finding that while most clerks reported that the Justices did not seem to have a particular pattern of deciding whether to make edits on their drafts, there is some evidence that Justices only substantially revised their draft opinions when particular issues were involved or when the cases were landmark or important cases).

37. The claim is often made that the discipline of writing out an argument is what enables the reader to test the logic of the argument. See Posner, supra note 23. Law clerks, who have been delegated the task of writing an opinion based on an argument that the judge has suggested, are perhaps less likely to second-guess the argument than the judge herself. Alternatively, some might think that law clerks are more likely to second-guess the logic of the argument and point out flaws. Once again, we do not know the answer. Knowing the information as to which opinions were solely authored versus which ones were not, however, might help us investigate further.
until we collect the information and run empirical tests on that information. Finally, assuming the information turns out to be useful, the authorship data might also be used as an outcome variable itself, and models could be constructed and tested to determine what factors determine—or at least correlate with—authorship rates.

5. Information for Law Clerks

Information about judicial writing proclivities will undoubtedly prove useful to law students applying for clerkships. The nature of a clerkship experience is a direct function of what the judge does. If the judge researches and writes all of her own opinions, there is little left for the clerks to do. If the judge is off skiing or at the beach for most of the year, only calling in on occasion, there will be an immense amount for the clerks to do and correspondingly little guidance from the judge.

Clerkships provide value to potential clerks in at least two ways: training and status. Law clerks are likely to receive minimal training if their judges are at the beach and have minimal contact with them. Clerks’ training is also likely to be minimal if their judges are so capable that they do all of their own research and writing. What the clerk who is seeking training wants, then, is the judge whose approach lies between the two extremes. Other clerks, who are seeking status and the power to shape the law, might want to look for a judge who delegates the major portion of opinion-writing discretion to the clerks. Clerks who primarily value leisure might prefer judges who need no help. Effective information about judges and their proclivities will assist prospective clerks in choosing judges who best suit their abilities and desires.

38. For example, clerkships with Justice Douglas were reputed to be among the worst because, at least in part, he did all his own writing and generally kept his clerks at a distance. See WARD & WEIDEN, supra note 3 (manuscript at ch. 5, at 8) (reporting on interviews with former clerks for Justice Douglas).

But will a more effective process for matching clerks to judges bring social benefits? In terms of efficiency, it should. Clerks who desire to have a greater input in the writing process will end up with judges willing to allow their clerks to do so. Likewise, clerks more interested in learning from a judge who actively manages opinion writing will be better able to find such judges. Whether such efficiency gains are great in magnitude and result in better judicial opinions, however, is unclear. To the extent the best clerks always opt for judges with a reputation as “feeder judges” for Supreme Court clerkships, information on authorship will not change the market much (unless it indirectly affects which judges are the feeder judges).

If clerks, in fact, prefer to work for judges who tend to author their own opinions or are otherwise more active in the judicial process, an indirect incentive effect on judges is possible. Judges would be forced to get more involved or accept lower-quality clerks. Alternatively, if judges’ reputations for providing inadequate training hurt their ability to get good law clerks, they might engage in more training activities for clerks.

6. Informational Benefits to Opinion Users

Knowing an author’s identity serves as a useful shortcut in a variety of settings. For example, in a bookstore, one might use authors’ names to decide which books to scrutinize and which ones to ignore. Similarly, when one conducts academic research, there are some authors who can be depended on for high-quality work and are therefore likely to be read first. Shortcuts such as an author’s identity are especially useful when the reader does not have the time or ability to read all of the available material and evaluate it piece by piece.

Lawyers and judges are faced with this dilemma on a daily basis. In attempting to construct their arguments, they have to draw from a vast body of available precedent and decide which arguments to use and which ones to reject or ignore. If there are multiple opinions that relate to a particular issue and there is but a limited amount of time to allocate to the task of analysis, a choice has to be made about which of those opinions should receive more attention. One of the variables used to sort through opinions is likely to be author identity. The dictates of formal precedent aside, opinions by judicial superstars like Hand, Friendly, and Cardozo will probably receive close scrutiny.

But what if it were known that while Judge Friendly wrote all his securities law opinions, he delegated his social security opinions to...
his law clerks? Or, what if Judge Easterbrook was known to write all his own opinions and Judge Posner was known to write none of his? Presumably, that would make a difference as well, with opinions authored solely by the judge receiving greater attention and those that were not receiving less attention. To push the argument further, it would help outsiders decide which opinions to concentrate on if the opinions were sorted based also on the various coauthors. So, if there were an opinion on the Second Circuit for which Judge X was the primary author, but to which Judge Hand had made a sufficient contribution to be named coauthor, the opinion might receive greater authority than if it were to appear under Judge X's name alone (assuming, for purposes of this point, that Judge X did not have a high reputation). Pushing even further, law clerks could be identified as coauthors or even primary authors when their contributions so warranted. The assumption would still be that the primary decision was made by the judge. But it would help outsiders who were attempting to determine how much weight to give a particular opinion to know who actually wrote the opinion.

Greater information on authorship may also add more credibility to—or alternatively quash—rumors on the supposed influence of law clerks in the writing process. There are rumors that Justice Thurgood Marshall wrote very few of his own opinions and that he wrote fewer and fewer as he advanced in age. There are rumors that Justice Thurgood Marshall wrote very few of his own opinions and that he wrote fewer and fewer as he advanced in age. There are rumors that Justice Thurgood Marshall wrote very few of his own opinions and that he wrote fewer and fewer as he advanced in age.

41. Justice Brandeis famously said that the reason why the Supreme Court's reputation was so high was that everyone knew that the Justices did all of their own work (a fact that is undoubtedly not true any longer). See William O. Douglas, The Court Years 1939-1975, at 172-73 (1980) (describing Justice Brandeis' views on the need for judges to author their own opinions). Judge Posner, in turn, explains:

The less that lawyers and especially other judges regard judicial opinions as authentic expressions of what the judges think, the less they will rely on judicial opinions for guidance and authority. . . . The more the thinking embodied in opinions is done by law clerks rather than by judges, the less authority opinions will have.


42. See Paul J. Wahlbeck et al., Ghostwriters on the Court?: A Stylistic Analysis of U.S. Supreme Court Opinion Drafts, 30 Am. Pol. Res. 166, 172 (2002) (describing the rumors regarding Justice Marshall's relative lack of involvement in the opinion-writing process). The public articulations of these rumors are contained in Bob Woodward & Scott Armstrong, The Brethren: Inside the Supreme Court 258 (1979), and Terry Eastland, While Justice Sleeps, Nat'l Rev., Apr. 21, 1989, at 24. See also Peter Huber, Advice to Justice Thomas, Forbes, Nov. 25, 1991, at 202 (finding that Justice Marshall's opinions during the 1990 term demonstrated four distinctive styles, corresponding to his four different law clerks). On the other side, those familiar with the workings of the Court and, in particular, Marshall's chambers, have disputed the view articulated above. See Juan Williams, Thurgood Marshall: American Revolutionary 370 (1998) (quoting both a close friend and former law clerk of Marshall); Mark Tushnet, Thurgood Marshall and the Brethren, 80 Geo. L.J. 2109, 2112 (1992) (stating that while Marshall may have relied "more heavily on his law clerks for opinion writing than did the other Justices during the early years of his tenure, . . . his practices were not wildly out of line with those of the others on the Court").
tice Kennedy’s opinion in Planned Parenthood v. Casey was the product of undue influence from one of his liberal clerks. And judges like William Douglas, Learned Hand, and Richard Posner are reputed to have written all their own opinions.

These rumors may unfairly hurt or benefit a judge’s reputation, effects that are especially problematic when some of the rumors are driven by stereotypes based on race and gender. Law clerk denials and confirmations occur, but these lack credibility since the law clerks have an incentive to do everything to heighten the reputation of their judge, because their own status is tied to the status of their judge. If, however, authorship can be determined using credible and verifiable methods, these rumors can be quashed or confirmed.

B. Downsides

When we began this project, we saw few disadvantages to collecting authorship information on judges. More information, we assumed, was a good thing. But a number of our colleagues, seeing our inquiry as problematic, disagreed. Their prime objection was that the venture was a waste of our time and resources because information about judicial authorship, even if it could be obtained, was useless—an objection that we hope Part II.A has answered. Below, we tackle two of the other objections we heard most often.

1. Danger of Unjustified Inferences

One could argue that an implicit message exists in just our attempt to test the degree of self-authorship among judges. Judges who do not author their own opinions will feel unfairly stigmatized, the argument goes, when it is not clear that they should. Judging is supposed to be about applying the law to the facts in an impartial and considerate manner. Whether the articulation of that application is

44. See Richard Lacayo, Inside the Court, TIME, July 13, 1992, at 29; Edward Lazarus, Disturb- ing Truths, Jurist: Books-on-Law (July 1998), at http://jurist.law.pitt.edu/lawbooks/revju98.htm. The claim that liberal law clerks have had excessive influence on the Justices was also famously made years earlier by Justice Rehnquist (prior to his joining the Court). See Wahlbeck et al., supra note 42, at 167. The controversy that ensued led one conservative Senator to ask for clerk confirmation by the Senate. Id.
46. See supra note 42 (citing statements by former Marshall clerks).
self-authored or not is a consideration of minimal value, at best.\footnote{47} The identity of the author does not matter, and for us to suggest otherwise by making this inquiry is disrespectful.

We disagree. One may not want judicial or author identity to be relevant, but there is reason to think that it is. Judges have individual preferences and respond to incentives. There is debate about the degree to which judges, as a specific subgroup, are driven by these factors, as opposed to other factors such as the norms of their profession or altruism toward society.\footnote{48} Given that debate, the necessary next step should be to test the robustness of the competing models, and that requires collecting information on variables such as authorship.\footnote{49} It may be that the empirical results will pleasantly surprise our critics and authorship identity will turn out not to matter. And if that is the case, there will not be any stigma. On the other hand, if self-authorship is related to productivity, quality, and other factors of importance, then a stigma should arguably fall on those judges who do not author their own opinions.

2. Imperfect Measurement

A second criticism takes aim at our methodology. Our critics point out that the types of statistical tests that we describe later in the Essay can at best produce an imperfect measure of authorship rates. Given that judges themselves have perfect information about their authorship levels, why not simply ask them about their practices? The point is fair. Judges do know more about their own practices, and we should try to ask them about what they do. But doing that alone is unlikely to be sufficient for two reasons. First, judges may choose simply not to respond to a survey instrument, especially if it is from some annoying law professor. Indeed, judges seemed irate when requested to respond to a survey sent out by the Senate’s Judi-
ciary Oversight Subcommittee.\textsuperscript{50} Second, there is a verifiability problem with judges self-reporting their practices. What’s to stop a judge from saying that she writes all her opinions (whether true or not)? Judges could ameliorate the verifiability problem somewhat by allowing their clerks free reign to disclose delegation practices. But we think this unlikely to happen.

Imperfect measures using computational techniques might be able to partially solve these problems by creating both an incentive to self-report and a verifiability mechanism. In other words, the revelation of partial information by an outside source may create an incentive for the possessor of the true information to fully reveal it, if only to correct the imperfect impression left from objective data on authorship.\textsuperscript{51} If the imperfect information is seen as having a high degree of reliability (for example, it is correct seventy percent of the time), then that will put pressure on judges who rank low on the imperfect measure, but whose true scores should be higher, to self-report their true practices. And to the extent there is a perceived verifiability problem with self-reporting, the judges may allow for verifiability through mechanisms like clerk disclosure.

III. TESTING AUTHORSHIP

So far, we have assumed that authorship identity could be determined. In theory it can be, but the task is not straightforward. Indeed, prior attempts at surveys on this very information have revealed little other than the fact that many judges do delegate some portion of the writing task to their clerks.\textsuperscript{52} In this Part we describe

\textsuperscript{50} See \textit{infra} note 52 (discussing the survey attempt). The Grassley survey may not be the best example, though. The judges’ annoyance at Senator Grassley’s attempt to survey them may have been more a product of the somewhat undiplomatic tone and context of that particular survey attempt. See \textit{infra} note 52.

\textsuperscript{51} For more on the argument that objective tournaments may help force participants to reveal information about themselves, see Scott Baker, Stephen Choi & Mitu Gulati, \textit{The Rat Race as an Information Forcing Device}, 81 Ind. L.J. (forthcoming 2006), \textit{available at} http://papers.ssrn.com/sol3/papers.cfm?abstract_id=649083.

\textsuperscript{52} There was a survey attempt directed by Senator Charles Grassley in 1996 as part of his work as chair of the Senate-Judiciary Subcommittee on Administrative Oversight and the Courts. See U.S. Senate-Judiciary Subcomm. on Admin. Oversight & the Courts, 104th Cong., Report on the January 1996 Judicial Survey (Part I, U.S. Courts of Appeal) 26-27 (Comm. Print 1996) [hereinafter Survey Report]. He sent a survey to all sitting federal judges that inquired about matters such as delegation to law clerks. See \textit{Now the Judges Face the Questions}, Legal Times, Feb. 5, 1996, at 8 (providing the full text of the Grassley questionnaire). The survey results suggested that most judges did not perceive an inappropriate level of delegation. See Survey Report, supra, at 26-27. (reporting that over seventy-five percent of the circuit court judges did not perceive a problem with the extent of delegation to law clerks); \textit{Oyez Surveys}, Legal Times, Aug. 12, 1996, at 3 (reporting that Grassley's "results show that most judges believe that law clerks do not play too large a role in judicial decision making"); Deborah Pines & Bill Alden, \textit{District, Circuit Judges Use Senate Survey to Boast, Gripe}, N.Y. L.J., Mar. 25, 1996, at 1, 4 (reporting that Judge Newman responded to the question of how much he delegated work to law clerks).
our preliminary attempts to determine authorship through other techniques.

The literature on the question of tracing authorship dates back to over a century ago. In recent years, these techniques have made it onto television talk shows and the front page of *The New York Times*. Stories have appeared about their application in the discovery of a new Shakespeare poem and in a variety of other high-profile settings, such as unmasking the anonymous author of *Primary Colors*. Law enforcement agencies and prosecutors have also used “forensic linguistics” in both the JonBenét Ramsey murder investigation and the Unabomber prosecution.

In attempting to answer these questions and others—such as whether Shakespeare wrote his plays or if they were really the work of Marlowe, Bacon, or some other contemporary—scholars have de-
developed a number of techniques.\textsuperscript{56} We make use of these techniques in discerning authorship for judicial opinions. Although some linguistics tools have found their way into FBI investigations and even into court,\textsuperscript{57} few legal academics have made any meaningful use of them in their research on judicial authorship.\textsuperscript{58} One reason for the lack of research is resource-related.\textsuperscript{59} In addition, determining authorship is difficult without a set of authentic texts for each judge (for example, texts where it is known to a certainty that the judge is the author) against which to compare the judicial opinions bearing the judge’s name.

There has been, however, at least one attempt to determine authorship in the judicial context. Paul Wahlbeck, James Spriggs, and Lee Sigelman used techniques from computational linguistics to determine the relative levels of delegation of the opinion-writing task to clerks by Justices Lewis Powell and Thurgood Marshall.\textsuperscript{60} We consider whether some version of that method could be used to determine authorship patterns for a larger sample of judges.

To get a sense of the difficulty of tackling this question and as a preliminary test of the effectiveness of standard authorship methodologies in determining judicial opinion authorship, we ran our authorship tests on a sample consisting of opinions for all the active federal circuit court judges for the period from January 1, 1998 to December 31, 2000—a total of three years worth of data.

Our tentative conclusion: the ranking task can be accomplished, but it requires a significant allocation of resources—at least more than the minimal amounts that law faculty typically use on their pro-

\textsuperscript{56} Inquiries into the authorship of Shakespearean texts have been multitudinous. For a discussion of some examples, see Barron Brainerd, The Computer in Statistical Studies of William Shakespeare, 4 COMPUTER STUD. HUMAN. & VERBAL BEHAV. 9 (1973), and C.B. Williams, Mendenhall’s Studies of Word-Length Distribution in the Works of Shakespeare and Bacon, 62 BIOMETRIKA 207 (1975). For more on inquiries into the true authorship of the texts we attribute to Shakespeare, see Ward Elliot, The Shakespeare Clinic, ELLIOT ONLINE, at http://govt.mckenna.edu/welliott/shakes.htm (last visited Feb. 5, 2005), and THE SHAKESPEARE AUTHORSHIP PAGE, at http://shakespeareauthorship.com (last visited Feb. 5, 2005).

\textsuperscript{57} See Foster, supra note 54, at 95-142 (describing the use of his techniques by both the FBI and the prosecutors in the Unabomber case); see also Bryan Niblett & Jillian Boreham, Cluster Analysis in Court, 1976 CRIM. L. REV. 175 (describing how the “cluster analysis” technique is used to verify criminal confession statements).

\textsuperscript{58} We should note though that at least two sets of legal scholars have discussed forensic linguistics in other contexts. See LAWRENCE M. SOLAN & PETER M. TIERSMA, SPEAKING OF CRIME: THE LANGUAGE OF CRIMINAL JUSTICE 149-80 (2005) (discussing the application of authorship attribution research to a wide variety of criminal investigations and cases); JOHN M. CONLEY & WILLIAM M. O’BARR, JUST WORDS: LAW, LANGUAGE, AND POWER 161-78 (2d ed. 2005) (discussing Donald Foster’s involvement in the JonBenét Ramsey murder investigation).

\textsuperscript{59} See infra text accompanying notes 86-88 (describing time-intensive nature of our authorship methodology).

\textsuperscript{60} See Wahlbeck et al., supra note 42, at 170-73.
jects. Moreover, significant “noise” may exist in comparing authorship of different judicial opinions that may make determining the true level of authorship for a particular set of opinions for any particular judge difficult. Nonetheless, the same reasons we offer in Part II.A of this Essay to support the value of determining authorship lead us to believe that more research is warranted in the area.

In Part III.A, we give a thumbnail sketch of the methodology used for doing authorship testing. In Part III.B, we then examine the question of how that methodology might be applied to the matter of ranking judges, and we provide results from generic tests of authorship drawn from computational linguistics. The generic tests perform poorly in distinguishing judges based on authorship. In Part III.C, we then report the results from better-tailored tests of authorship for the judicial context. Tests that control for the subject matter of judicial opinions perform better in assessing judicial authorship.

A. Tracking Judicial Fingerprints

The basic proposition here is that writers have styles of their own. Just as all of us have our own styles of walking, talking, singing, shooting a photograph or movie, throwing a baseball, and playing a guitar, we also have particular writing styles. Some of us have styles that are more distinctive than others. This is especially true when we are experts in a field and, as a result, are extensively active within the field. Michael Jordan’s style of shooting a basketball, Serena Williams’ style of serving a tennis ball, and Sachin Tendulkar’s style of hitting a cricket ball are so distinctive that even casual fans are likely to identify the players from just their playing styles. Our more literate friends can readily recognize passages from authors such as Jane Austen, Ernest Hemingway, and F. Scott Fitzgerald.

Testing authorship by way of an author’s idiosyncratic style requires distilling the basic elements of the style that differ from that of other authors in order that these basic elements can be used as identifiers. For example, if one knows that author A is partial to using the word “hath” and intensely dislikes the word “have,” one might think it unlikely a document that has zero uses of “hath” and multiple uses of “have” belongs to him. The information about author A’s distinctive style characteristics is combined with information about the patterns in the document of unknown authorship, and Bayes’ Theorem is used to determine the probability that A authored the mystery text.61

61. The standard technique to make these probability calculations is to use Bayes’ Theorem. One of the problems with the use of the naïve Bayes model that has been pointed out is that the calculation assumes randomness (and, of course, words in a text are a function of each other as opposed to completely independent and random). Nevertheless, despite the
The problem with this technique is that one has to identify a large set of unique identifiers (such as uses of “hath” in place of “have”). Without such a set of identifiers, it becomes difficult to distinguish authors from one another. Relying on unique stylistic markers also requires expertise in identifying the author’s style and will work only for authors who have developed identifiable traits in the first place. This narrows considerably the usefulness of such techniques. Not all authors use unique stylistic markers in their writing (Choi and Gulati, for example, are quite mundane in their styles, although they are both partial to text parentheticals).

More broadly applicable techniques for authorship determination build on the same basic idea but look to a variety of style patterns rather than particular identifiers. Here, the premise is that authors’ writings follow patterns. These patterns may be, in part, a function of stylistic preferences, such as the use of the word “hath.” Patterns may also consist of the use of a series of more common words and the order in which such words are used. Rather than look for specific styles, authorship may be determined through an examination of the overall frequency of specific words and patterns of words throughout the entire text.

Word choice patterns (that is, diction) are also likely to be a function of bounded rationality. People engaged in the task of writing, just as with other tasks, are constrained by their cognitive capacities. There is likely to be a finite set of words that are part of a single individual’s active vocabulary. One uses these words more often than other words. Other words, though part of one’s vocabulary, are used only rarely in one’s own speech or writing yet are immediately recognizable in both the text and speech of others. And then there will be words the author simply does not know. Individual authors are likely to exhibit distinct patterns of noun and verb usage, distinct patterns

for combining words, and distinct patterns of starting and ending sentences. All of these patterns can translate into frequency of use. Given a set of authentic texts for any author, a full set of these frequencies can be calculated. These frequencies can then be compared to the frequency patterns for the document whose authorship is unknown to determine whether the same author wrote the document in question.62

Among the first serious treatments of the mathematics of word distributions, to our knowledge, was that by G. Udny Yule in 1938, in a study of *The Imitation of Christ*.63 The defining studies in the area, however, are four studies published in the 1960s: Louis Milic’s tests of Jonathan Swift’s prose,64 A.Q. Morton and James McLeman’s study of the Pauline Epistles,65 Alvar Ellegård’s study of the *Junius Letters*,66 and Frederick Mosteller and David Wallace’s work on *The Federalist* papers.67 The success of these projects and improvements of computer technology spurred an expansion of the literature on style statistics in the decades since. For interested scholars, extended treatments and a steady stream of sophisticated journal articles exist.68

---


63. See Farringdon, *supra* note 61, at 549-50; see also G. Udny Yule, *The Statistical Study of Literary Vocabulary* (1944) (investigating the measurement of word distribution). Prior to Yule, there is the suggestion that stylistic habits might be used to detect authorship in a letter from the mathematician Augustus De Morgan in 1851. See Farringdon, *supra* note 61, at 549 (citing SOPHIE ELIZABETH DE MORGAN, MEMOIR OF AUGUSTUS DE MORGAN 215-16 (1882)). There is also the work on frequency distributions of words by Zipf in 1932. See GEORGE KINGSLEY ZIPF, SELECTED STUDIES OF THE PRINCIPLE OF RELATIVE FREQUENCY IN LANGUAGE (1932).

64. LOUIS TONKO MILIC, A QUANTITATIVE APPROACH TO THE STYLE OF JONATHAN SWIFT (1967).


68. Among the extended treatments or collections of articles are J.F. BURROWS, COMPUTATION INTO CRITICISM: A STUDY OF JANE AUSTEN’S NOVELS AND AN EXPERIMENT IN METHOD (1987); THE COMPUTER IN LITERARY AND LINGUISTIC STUDIES (Alan Jones & R.F. Churchhouse eds., 1976); PAULA R. FELDMAN & BUFORD NORMAN, THE WORDWORTHY COMPUTER: CLASSROOM AND RESEARCH APPLICATIONS IN LANGUAGE AND LITERATURE (1987); ANTHONY KENNY, THE COMPUTATION OF STYLE: AN INTRODUCTION TO STATISTICS FOR STUDENTS OF LITERATURE AND HUMANITIES (1982); LITERARY COMPUTING AND LITERARY CRITICISM: THEORETICAL AND PRACTICAL ESSAYS ON THEME AND RHETORIC (Rosanne G. Potter ed., 1989); and A.Q. MORTON, LITERARY DETECTION: HOW TO PROVE AUTHORSHIP AND FRAUD IN LITERATURE AND DOCUMENTS (1978). Current articles in the field can be found in the journals Literary and Linguistic Computing and Computers and
The question, then, is whether these techniques can be applied to judicial opinions. The answer is not obviously in the affirmative.

The fact that Jane Austen’s or Ernest Hemmingway’s texts have distinctive styles and are easily recognizable does not mean that the writing styles of more ordinary mortals will be recognizable. Moreover, testing judicial authorship is potentially even more problematic because of the institutionalized nature of judicial writing. To the extent that judges consciously try to follow some institutional style (and, therefore, consciously suppress their own style), the difficulties in identifying particular judicial authorship are likely further exacerbated. Our project, however, is to identify the degree of judge authorship versus clerk authorship.

There is reason to believe that judicial style is likely to be different from clerk style. Judges, who tend to have been experienced lawyers or academics earlier in their careers, are likely to be more confident in their writing than their clerks, who tend to be fresh out of law schools. Judges, because of their high level of skill and confidence, may write shorter opinions with fewer citations and footnotes. Judges also are likely to attack the central issue in their cases directly. Clerks, by contrast, because of their inexperience, may tend to write lengthy opinions with numerous citations and footnotes.

As noted earlier, the sole application of the linguistic techniques discussed above to the question of judicial delegation to law clerks is the paper by Wahlbeck, Spriggs, and Sigelman. Court lore has it that Justice Thurgood Marshall delegated writing tasks heavily to his law clerks, whereas Justice Lewis Powell was more actively involved in the production of opinions. Wahlbeck and his coauthors attempted to use information available from the Justices’ papers.

---

the Humanities. See Wahlbeck et al., supra note 42, at 189 n.14 (suggesting these two journals and a number of the volumes mentioned above for background in statistical stylistics).


70. The point that overreliance on law clerks can significantly influence the style of opinions has been made by a number of commentators. See COHEN, supra note 3, at 94 (making the point that delegation of opinion writing to law clerks can affect the clarity and style of the opinion); POSNER, supra note 41, at 115 (suggesting that the excessive numbers of footnotes, citations, and words in opinions are all a product of heightened levels of delegation to law clerks); Wahlbeck et al., supra note 42, at 173 (making the point that clerks are more likely than judges to rely extensively on multifactor and balancing tests in the opinions that they draft) (citing ANTHONY T. KRONMAN, THE LOST LAWYER (1993), and POSNER, supra note 3). But see Samuel Estreicher, Conserving the Federal Judiciary for a Conservative Agenda?, 84 MICH. L. REV. 569, 574 (1986) (agreeing that law clerks may be partially responsible for the excessive numbers of footnotes and the deadening style of current opinions but also stating that part of the blame may also lie with computerization).

71. See Wahlbeck et al., supra note 42.

72. See id. at 170-72; see also sources cited supra note 42.
containing clerk identifiers on bench memoranda and draft opinions, to see whether they could detect the “fingerprints” of the clerks in the opinions with which the two Justices were involved. The greater the involvement of the Justice in the actual drafting, they hypothesized, the less likely it would be that the fingerprints of the clerks would be detectable. In order to perform the detection task, they used eight different frequency measures. These were average footnote length, average sentence length, average word length, word length diversity, sentence length diversity, footnote frequency, type-token ratio, and the once-word rate. Consistent with their initial hypothesis, Wahlbeck, Spriggs, and Sigelman found that the fingerprints of judicial clerks are clearer in Marshall’s opinions than in Powell’s opinions. This implied that Powell had a greater hand in authoring his own judicial opinions.

We attempt to adapt the authorship methodology used in computational linguistics to the task of ranking authorship rates for judges. We note, however, several important points. Ideally, we could perform the following two step methodology:

**Step 1:** Run the authorship methodology on a set of authentic writing samples from a particular judge to serve as the baseline of comparison.

**Step 2:** Compare the judge’s judicial opinions against the authentic writing samples to determine whether the judge in fact authored the opinions. The greater the discrepancy between the opinions and the authentic writing samples, the less likely the probability that the particular judge authored the opinions.

For most judges, there is unlikely to be an available set of authentic samples of the judge’s own writing that can be used to calculate baseline frequencies. We therefore are unable to determine how many of a particular judge’s opinions the judge actually wrote.

Nonetheless, at least two uses for the authorship methodology are possible in judge-related authorship studies. First, we can make a comparison of a subset of a particular judge’s opinions to each other.

---

73. Wahlbeck et al., supra note 42, at 168, 178-82.
74. See id. at 174.
75. Id. at 176-77.
76. Id. Type-token ratio is “the number of different words in an opinion (types) as a percentage of the total number of words in the opinion (tokens).” Id. at 176. Once-word refers to “the relative frequency of words that appear exactly once in an opinion.” Id.
77. Id. at 182-83.
78. For a subset of judges such as former law professors, we can find articles that they authored. There are two problems with using these articles as the basis for opinion testing. First, article writing is a different genre from opinion writing; and second, law review editors are involved in article writing. That said, since it is these same law review editors who often become law clerks, maybe opinion styles and law review article styles are more similar than we think.
Assuming that consistency in the sentence patterns and other aspects of an author’s style is positively correlated with authentic authorship, one can then rank the judges relatively against one another. If the judge authored all the opinions, presumably the opinions will receive a high same-authorship score. If the judge did not author all of the opinions, then the set of opinions will receive a low same-authorship score.

This variation measure across a set of opinions for which a particular judge is named as the author, it should be cautioned, will provide a high score for a judge who delegates all her work to a single, permanent law clerk.79 It would also likely give higher scores to judges who imposed their styles on the writing process by doing heavy editing as opposed to doing any of the actual writing. What we would be measuring, then, is best described as the relative degree of various judges’ involvement in the writing process.

Second, we can use the authorship methodology to make relative comparisons of how authorship for a particular judge has changed over time. We may not know if Judge A authors her opinions or not. But we can see if the level of authorship (regardless of the starting point) has declined or increased over time. Consider Supreme Court Justices. One hypothesis is that as individual Justices age, they play a diminishing role in writing their own opinions or, alternatively, in monitoring the work product of their clerks.80 We could compare the authorship score for a particular Justice when in her fifties against the authorship score when in her eighties. If the hypothesis is correct, one would expect to see a relative drop in the same-authorship score.

In this Essay we discuss only the first application of the authorship methodology to ranking judges (the judge versus judge comparison). Doing a full-scale ranking of any meaningful sample of judges is beyond the scope of this project—and our research budgets. We do, however, briefly report the results of some preliminary tests and discuss the limitations of our methodology as well as possible adjustments to improve on our work. We leave for another paper the comparison of judges relative to their past selves.81

79. Such a result is unlikely, nonetheless, where the sample of opinions for a judge are drawn across multiple years (and thus involve multiple different sets of clerks).
80. See WARD & WEIDEN, supra note 3 (manuscript at ch. 5, at 11) (reporting evidence on how the authorship patterns of Justices such as Rehnquist and Blackmun changed over their time on the Court, with both Justices delegating more of the opinion writing to clerks as they gained seniority).
81. We are in the process of collecting these data now.
B. Ranking Judges

To test authorship rates, we start with a set of generic tests commonly used in other studies of authorship. We do not control for different types of documents in the generic tests but take instead a randomly selected set of opinions for a judge and compare them to one another. We label these generic tests our “black-box” tests because of the lack of subject matter controls.

The two black-box tests that we use are recognized authorship-testing methods—only two out of a variety of tests available. We describe them briefly before discussing the results of the tests. We rely primarily on the GZip compression technique, which compresses documents based on the similarities in the basic linguistic building blocks of the two files. The GZip algorithm looks for repeated phrases within a threshold of the last 8000 characters of text analyzed (a number that can be made larger, depending on the specific program). The longer and more frequent the repeated phrases, the higher the score the algorithm accords to the text; the end result is greater compression.

---


83. In using and applying the GZip method, we relied on the computer scientists at In-App software company in Trivandrum, India—Satish Babu, M.C. Jayakrishnan, and R.V. Suchithra—who consulted with colleagues at the computer science department at Kerala University to develop a simple application of the GZip method for our project. Should readers be interested in more details on the GZip application, please email questions to: Satish Babu, sb@inapp.com; M.C. Jayakrishnan, jayan@inapp.com; or R.V. Suchitra, suchi@inapp.com.

Our use of the compression method to test authorship is simplistic. For discussions of far more sophisticated applications of the compression methodology to test for authorship that are beyond the mathematics and statistics skills of the two authors here, see, for example, Dario Benedetto et al., Language Trees and Zipping, 88 PHYSICAL REV. LETTERS 048702-1 (2002); Eibe Frank et al., Text Categorization Using Compression Models, in PROCEEDINGS: DCC 2000 DATA COMPRESSION CONFERENCE 555 (James A. Storer & Martin Cohn eds., 2000); D.V. Khmelev, Disputed Authorship Resolution Through Using Relative Empirical Entropy for Markov Chains of Letters in Human Language Texts, 7 J. QUANTITATIVE LINGUISTICS 201 (2000); O.V. Bukhshchina et al., Using Literal and Grammatical Statistics for Authorship Attribution, 37 PROBS. INFO. TRANSMISSION 172 (2001); and William J. Teahan & David J. Harper, Using Compression-Based Language Models for Text Categorization (2001), available at http://citeseer.ist.psu.edu/teahan01using.html (last visited Feb. 11, 2005). Simphile, a program from Geneffects, is a commercial implementation of author recognition, using the same technique. The methodology has also reportedly been used for gene sequence matching in bioinformatics and authorship of music and art. For more information, see GENEFECTS, SIMPHILE, at http://www.geneffects.com/simphile (last visited Feb. 8, 2005).
Compression serves as a metric for redundancy or “entropy” and, as a result, is a metric for comparing authorship. The compression for a combined file containing an English document combined with an Italian one, for example, will be significantly less than that of a combined file containing two similar length documents in English. Along those lines, research suggests that the compression of files from the same authors will be significantly greater than that of files from different authors (due to the greater number of repeated phrases).

We can calculate a compression score for any two documents (labeled A and B) as follows:

\[
\text{Compression Score} = \frac{P}{Q}
\]

where:

\[\text{Size}(\cdot) = \text{The size in bytes of the compressed file}\]
\[P = \text{Size}(\text{Document A}) + \text{Size}(\text{Document B})\]
\[Q = \text{Size}(\text{Documents A combined with B})\]

\(P\) is always greater than \(Q\). If documents A and B are written by the same author, we assert that \(P\) will be much greater than \(Q\). Put another way, if documents A and B display many similarities, compressing the combination of the two documents will result in significant space savings when compared with the compression of each document separately. On the other hand, if the documents are written by different authors, \(P\) will, in the extreme, approach \(Q\) (as the compression program will find fewer common phrases between the documents). If documents A and B are completely different, then combining the two will result in no additional compression.

For each of the ninety-eight judges in our sample, we collected four text samples of 8000 characters each. Text samples were chosen at random from opinions at least 32,000 characters in length that were written between 1998 and 2000. In our black-box tests, we did not control for the subject matter of the opinion. A particular judge may have four criminal law opinions or, alternatively, a set of crim-
nal, constitutional, commercial, and securities-regulatory opinions. As we discuss below, however, this lack of control for the type of opinion may introduce genre-specific “noise” in our results. The compression score for four criminal law opinions from a judge who does not write her own opinions may be greater (leading to a higher same-authorship score) than the score for four opinions from different areas of the law for a judge who does in fact write her own opinions.

Since the goal was to test authorship style, we eliminated all portions of the documents that were not a product of this style. That meant removing all West headnotes, citations, and quotes from each document. Commonly used words such as “and” and “the” were also eliminated. Cleaning the documents was the most time-consuming task in this project because it had to be done manually. It took us 500-plus hours of research-assistant time to clean the pieces of text that satisfied our conditions for each of the ninety-eight judges in our sample. Cleaning the documents is not strictly required but does reduce the level of noise in the same-authorship score results, since, for example, two judges with dramatically different levels of “true” authorship may receive the same GZip score if the compression of the words “and” and “the” swamp all other differences between the documents.

We then made pairwise comparisons of these four pieces of text (labeled $A_1$, $A_2$, $A_3$, and $A_4$ below) and generated a four-by-four matrix for each judge with $P/Q$ scores for each of the pairs:

<table>
<thead>
<tr>
<th></th>
<th>$A_1, A_1$</th>
<th>$A_1, A_2$</th>
<th>$A_1, A_3$</th>
<th>$A_1, A_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_2, A_1$</td>
<td>$A_2, A_2$</td>
<td>$A_2, A_3$</td>
<td>$A_2, A_4$</td>
<td></td>
</tr>
<tr>
<td>$A_3, A_1$</td>
<td>$A_3, A_2$</td>
<td>$A_3, A_3$</td>
<td>$A_3, A_4$</td>
<td></td>
</tr>
<tr>
<td>$A_4, A_1$</td>
<td>$A_4, A_2$</td>
<td>$A_4, A_3$</td>
<td>$A_4, A_4$</td>
<td></td>
</tr>
</tbody>
</table>

Our goal was to measure the degree of variation in the writing style (here, measured by compression levels as determined using the GZip program). One measure of this variance is provided by the eigenvalue of the matrix. The higher the eigenvalue, the higher the probability of sole authorship.

---

87. This is not to suggest that these commonly used words might not be useful with a different type of authorship test. One could, for example, run frequency tests on the relative uses of these common words, because different authors likely have different styles with respect to these words as well (some authors likely use more of them than others). One rationale for using measures of these “function” words (such as “and,” “the,” and “of”) is that the rate of their use is a function of the unconscious or habitual element of writing and, therefore, a better indicator of true authorship. For discussions, see J.F. Burrows, Word-Patterns and Story-Shapes: The Statistical Analysis of Narrative Style, 2 LITERARY & LINGUISTIC COMPUTING 61 (1987), and David L. Hoover, Statistical Stylistics and Authorship Attribution: An Empirical Investigation, 16 LITERARY & LINGUISTIC COMPUTING 421 (2001).

88. The four-by-four matrix of compression scores for each judge’s four opinions is symmetric, since $P/Q$ will be the same for the pairs $[A_1, A_2]$ and $[A_2, A_1]$. By definition, a symmetric square matrix possesses $n$ real eigenvalues (where $n$ is the order of the ma-
To test the effectiveness of the GZip methodology for determining judicial authorship, we identified a set of judges who, a priori, are known to write their own opinions—our “test” judges. An informal survey of more than two dozen judges and law clerks, who requested that we not name them, produced a high degree of agreement on the names of three federal circuit court judges who regularly author their opinions: Richard Posner, Frank Easterbrook, and Michael Boudin. More specifically, we were told that Posner drafts every word of his opinions; that Easterbrook may allow his clerks to draft one or two opinions a year but drafts the remainder himself; and that Boudin uses a combination of very heavy editing of clerk drafts and self-authorship. To reiterate, all three are reputed to fall on the high end of the scale of self-authorship. Assuming this informal information to be correct, the scores for these judges should be among the highest. If the GZip methodology is effective in identifying judges who author their own opinions, our three test judges should rank highly among our sample of ninety-eight judges. We do not perform the alternative test of identifying several judges reputed not to write their own opinions and assessing whether the GZip methodology succeeds in assigning a poor ranking to such judges. Unfortunately, for purposes of our tests, the judges and clerks we surveyed were reluctant to identify those least inclined to author their opinions.

We report in Table 1 below the ranking of the top fifteen judges out of our sample of ninety-eight judges based on the eigenvalue score from our GZip test. Higher eigenvalues correlate with a greater likelihood of self-authorship. In addition, we report the ranking of our three test judges.

Eigenvalues have several applications. In this context, we use it as a single metric that measures the variability among the elements of the matrix. We borrowed the reasoning for using eigenvalues from factor analysis, where they are used to measure variability.

89. A fuller study would need to use additional controls, particularly if there is doubt about whether Posner, Easterbrook, and Boudin actually author the major portion of their own opinions. Two possible other controls are checking whether the authorship tests return higher scores for (a) dissents than for majority opinions, because informal information suggests that dissents are more personal to the judge and, we assume, more likely to be self-authored and (b) judges in the distant past than for judges of today, because judges in the distant past are reputed to have written a far larger fraction of their own opinions since they did not face the caseload pressures of current judges. We did not use either of these two controls because of inadequacies in our dataset. On the points about dissents being more personal and judges in the past writing more of their own opinions, see Blomquist, supra note 6, at 86-92 (discussing the characteristics of dissent styles); Richman & Reynolds, supra note 8, at 278-79 (discussing how the increases in caseloads from the time of Learned Hand to the present have resulted in increased levels of delegation of the opinion-writing task); Frankel, supra note 3 (bemoaning the passing of the Leaned Hand style of opinion writing, where there was little delegation of the writing task to the law clerks); cf. Wahlbeck et al., supra note 42, at 168 (pointing out that the conventional view that judges in the past single-handedly crafted all their own opinions whereas today’s judges lean heavily on their clerks is an oversimplification, because there is evidence suggesting that clerks have long played a significant role in the opinion-drafting process).
Somewhat surprisingly (and to our dismay after expending 500 hours of resources), none of our three test judges show up as top scorers on the GZip tests. Boudin, in fact, scores near the very bottom of the judges, and only Easterbrook is in the top half of judges.

We provide one variation on the GZip methodology. Instead of computing the eigenvalue of the four-by-four matrix of GZip scores, we classify the compression score for any two opinions for a particular judge as either high, middle, or low.90 We then look at the difference between the number of high compression scores (high likelihood of sole authorship) and the number of low compression scores (low

---

90. We take the lower triangular numbers of the four-by-four matrix (the upper triangular numbers can be ignored since the matrix is symmetric, while the diagonal can be ignored since it represents a self-referential relation). The overall range of the $P/Q$ compression numbers across all judges goes from 3.0 to 6.0. We classify each score in the lower triangular portion of the matrix into one of three qualitative groups as follows:

<table>
<thead>
<tr>
<th>Low</th>
<th>3.0 – 4.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium</td>
<td>4.2 – 4.8</td>
</tr>
<tr>
<td>High</td>
<td>4.8 – 6.0</td>
</tr>
</tbody>
</table>

---

**Table 1**

**RANKING BASED ON GZip METHODOLOGY**

<table>
<thead>
<tr>
<th>RANK</th>
<th>JUDGE</th>
<th>CIRCUIT</th>
<th>OPINIONS EIGENVALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Randolph, A. Raymond</td>
<td>DC</td>
<td>169.01</td>
</tr>
<tr>
<td>2</td>
<td>Ginsburg, Douglas H.</td>
<td>DC</td>
<td>160.29</td>
</tr>
<tr>
<td>3</td>
<td>Manion, Daniel A.</td>
<td>7</td>
<td>150.58</td>
</tr>
<tr>
<td>4</td>
<td>Garza, Emilio M.</td>
<td>5</td>
<td>145.91</td>
</tr>
<tr>
<td>5</td>
<td>McKee, Theodore A.</td>
<td>3</td>
<td>140.23</td>
</tr>
<tr>
<td>6</td>
<td>Moore, Karen Nelson</td>
<td>6</td>
<td>139.44</td>
</tr>
<tr>
<td>7</td>
<td>Tjoflat, Gerald Bard</td>
<td>11</td>
<td>137.95</td>
</tr>
<tr>
<td>8</td>
<td>Batchelder, Alice M.</td>
<td>6</td>
<td>137.46</td>
</tr>
<tr>
<td>9</td>
<td>Calabresi, Guido</td>
<td>2</td>
<td>137.03</td>
</tr>
<tr>
<td>10</td>
<td>Flaum, Joel M.</td>
<td>7</td>
<td>135.83</td>
</tr>
<tr>
<td>11</td>
<td>Walker, John M., Jr.</td>
<td>2</td>
<td>135.00</td>
</tr>
<tr>
<td>12</td>
<td>Gilman, Ronald Lee</td>
<td>6</td>
<td>134.86</td>
</tr>
<tr>
<td>13</td>
<td>Dennis, James L.</td>
<td>5</td>
<td>134.82</td>
</tr>
<tr>
<td>14</td>
<td>DeMoss, Harald R., Jr.</td>
<td>5</td>
<td>134.57</td>
</tr>
<tr>
<td>15</td>
<td>Ebel, David M.</td>
<td>10</td>
<td>134.48</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>44</td>
<td>Easterbrook, Frank H.</td>
<td>7</td>
<td>130.19</td>
</tr>
<tr>
<td>50</td>
<td>Posner, Richard A.</td>
<td>7</td>
<td>129.89</td>
</tr>
<tr>
<td>97</td>
<td>Boudin, Michael</td>
<td>1</td>
<td>121.81</td>
</tr>
</tbody>
</table>
likelihood of sole authorship) while ignoring the middle scores. We predict that judges who self-author their opinions will tend to receive a much greater number of high-compression-score opinions, after netting out all the low-compression-score opinions. Table 2 reports the rankings based on this GZip variation.

**Table 2**  
**Ranking Based on Number of High - Low GZip Scores**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Judge</th>
<th>Circuit</th>
<th>Number of High Compression Scores Minus Number of Low Compression Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Randolph, A. Raymond</td>
<td>DC</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>Manion, Daniel A.</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Ginsburg, Douglas H.</td>
<td>DC</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Garza, Emilio M.</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>McKee, Theodore A.</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Cole, R. Guy, Jr.</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Gilman, Ronald Lee</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Sentelle, David B.</td>
<td>DC</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Tashima, A. Wallace</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>Ebel, David M.</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>Tjoflat, Gerald Bard</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>Posner, Richard A.</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>Easterbrook, Frank H.</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>O'Scannlain, Diarmuid F.</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>Rovner, Ilana Diamond</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>98</td>
<td>Boudin, Michael</td>
<td>1</td>
<td>0.5</td>
</tr>
</tbody>
</table>

As with the initial GZip methodology, our test judges fail to score consistently well in the GZip score variation. Posner scores in the top fifteen, coming in tied at number ten with three other judges. Easterbrook also scores highly, tied at thirteenth. Forty-seven other judges, however, are also tied at thirteenth. Boudin is the lowest-ranked judge in terms of authorship. Given the small number of each judge's opinions that we examined, differences among judges in the GZip variation are not great enough to make fine-tuned distinctions among judges. Our first cut at both black-box tests resulted in failure: Neither GZip test produced results consistent with our a priori information that Posner, Easterbrook, and Boudin author their own opinions more than other circuit court judges.
What does the foregoing suggest? First, it may be that the small sample of data that we used for each judge, four randomly selected opinions, was inadequate. We may need to use a larger set of opinions for each judge for the black-box tests to have any traction. Moreover, our opinions are all selected within a narrow time frame (1998 to 2000). A high authorship score for a particular judge may indicate that the same clerk wrote some, if not all, of the four opinions for the judge. Choosing opinions spread out across a longer time frame will reduce the possibility that a high authorship score is due, in fact, to a particular clerk.

Second, the fact is that legal writing in a particular subject matter area contains genre-specific language. Tests such as the GZip algorithm look for similar word patterns. Legal opinion writing is likely to have considerable genre-specific commonality. The vast majority of opinions are likely to mention phrases such as “standard of review” or “summary judgment” or “motion to dismiss.” If there are enough of these common phrases in all the opinions, they may swamp the calculations. Put differently, legal opinion writing as a genre may have such a distinct style that genre-specific tests must be devised for it.

C. White-Box Tests

Our simple black-box tests without subject matter controls failed to distinguish judges based on the degree of authorship. While the black-box tests worked well in comparing generic texts with one another, we suspect that they failed in the judicial context due to the specialized nature of many types of judicial opinions. Opinions of the same genre, for example, may use various forms of jargon and other common phrases that are shared in the opinions of different judges. Even judges who do not self-author their own opinions will receive a high same-authorship score for opinions within the same genre.

Despite the failure of our simple black-box tests, we contend that additional tests geared to controlling for the subject matter of specific opinions may still work to distinguish judges based on the degree of opinion authorship. We call these tests our “white-box” tests. At least three categories of white-box tests are possible based on citation practices, language patterns, and a more nuanced version of the black-box GZip methodologies.

1. Citation Practices

Citations are a key element in the crafting of judicial opinions. Judges who write their own opinions may display specific patterns in the opinions they cite, tending to cite opinions with which they are more familiar. We perhaps acted too hastily in cleaning the opinions used in our black-box tests of the citations to other opinions. The pat-
terns of citations themselves provide crucial information on authorship. Other things equal, two opinions with the same citation patterns (for example, citing the same set of opinions) are more likely to be authored by the same author than two opinions without the same citation pattern.

One type of citation pattern involves self-citations. Landes, Lessig, and Solimine conjecture that judges who write their own opinions are likely to be more familiar with these opinions, which would lead to greater self-citation rates.\textsuperscript{91} Law clerks, by contrast, are likely to simply conduct Westlaw searches or draw from the parties’ briefs. A higher average number of self-citations per opinion is therefore likely to correlate with a greater likelihood of authorship.

Importantly, self-citation patterns are at least somewhat invariant with the subject matter of an opinion. A judge who tends to cite her own work will cite her own criminal law cases when writing a subsequent criminal law opinion and her own securities law cases when writing a subsequent securities law opinion. Subject matter may, nonetheless, still be important if a judge writes opinions in a particular area more frequently because she will, as a result, cite her own opinions in that area more frequently.

As a quick test of the hypothesis that self-citation rates correlate with authorship, we examine the self-citation patterns using our 1998 to 2000 dataset. We focus in particular on whether the three test judges reputed to author their own opinions are among the top scorers on self-citations. Table 3, infra, reports the self-citation rates ranking for our test judges and the top fifteen scorers in our ranking. The self-citation rate is defined as the average number of self-citations (measured from 1998 to 2003) to each judge’s opinions written in the 1998 to 2000 sample time period. Because we focus on each judge’s opinions from the same three-year period, each judge starts with a similar pool of opinions that they may self-cite. In addition, the use of a pool of opinions from the same time period controls for changes in self-citation patterns over time.

\textsuperscript{91} See Landes et al., supra note 14, at 274 (“It is not implausible that judges who write their own opinions will cite themselves more frequently than judges who do not—if only because they have a greater familiarity with their own prior opinions.”).
### Table 3

**Ranking Based on Self-Citation Rate**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Judge</th>
<th>Circuit</th>
<th>Average Self-Citations to Each Opinion</th>
<th>T-Test of Difference with the Median Judge* (Equal Variances)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Selya, Bruce M.</td>
<td>1</td>
<td>2.44</td>
<td>7.00**</td>
</tr>
<tr>
<td>2</td>
<td>Wollman, Roger L.</td>
<td>8</td>
<td>1.56</td>
<td>3.93**</td>
</tr>
<tr>
<td>3</td>
<td>Lynch, Sandra L.</td>
<td>1</td>
<td>1.48</td>
<td>4.15**</td>
</tr>
<tr>
<td>4</td>
<td>Posner, Richard A.</td>
<td>7</td>
<td>1.46</td>
<td>2.81**</td>
</tr>
<tr>
<td>5</td>
<td>Clay, Eric L.</td>
<td>6</td>
<td>1.40</td>
<td>4.44**</td>
</tr>
<tr>
<td>6</td>
<td>Carnes, Ed</td>
<td>11</td>
<td>1.31</td>
<td>4.02**</td>
</tr>
<tr>
<td>7</td>
<td>Garland, Merrick B.</td>
<td>DC</td>
<td>1.31</td>
<td>4.52**</td>
</tr>
<tr>
<td>8</td>
<td>Moore, Karen Nelson</td>
<td>6</td>
<td>1.23</td>
<td>3.12**</td>
</tr>
<tr>
<td>9</td>
<td>Kelly, Paul J., Jr.</td>
<td>10</td>
<td>1.12</td>
<td>3.13**</td>
</tr>
<tr>
<td>10</td>
<td>Easterbrook, Frank H.</td>
<td>7</td>
<td>1.10</td>
<td>2.04**</td>
</tr>
<tr>
<td>11</td>
<td>Ebel, David M.</td>
<td>10</td>
<td>1.08</td>
<td>2.73**</td>
</tr>
<tr>
<td>12</td>
<td>Coffey, John L.</td>
<td>7</td>
<td>1.03</td>
<td>2.15**</td>
</tr>
<tr>
<td>13</td>
<td>Murphy, Michael R.</td>
<td>10</td>
<td>1.03</td>
<td>2.61**</td>
</tr>
<tr>
<td>14</td>
<td>Kanne, Michael S.</td>
<td>7</td>
<td>1.02</td>
<td>2.07**</td>
</tr>
<tr>
<td>15</td>
<td>Marcus, Stanley</td>
<td>11</td>
<td>1.02</td>
<td>3.21**</td>
</tr>
<tr>
<td></td>
<td>Boudin, Michael</td>
<td>1</td>
<td>0.61</td>
<td>0.89</td>
</tr>
</tbody>
</table>

* The median judge (Mary Beck Briscoe) is chosen as the forty-ninth ranked judge out of ninety-eight total judges.

** Indicates a significance level of 5%. The self-citation rate for Mary Beck Briscoe was 0.35 to each opinion.

In Table 3, we observe that both Posner and Easterbrook are in the top ten judges out of the sample of ninety-eight circuit court judges in terms of self-citation rates. Boudin, on the other hand, is ranked twenty-seventh, although he still is in the top half of all judges. For each judge, we perform a two-sided t-test assuming equal variances between that judge’s self-citation rate and the median judge’s—Mary Beck Briscoe—self-citation rate. All the top fifteen judges are significantly different at the five percent confidence level from the median judge.

Note that the t-statistic test we perform in comparison with the median judge does not tell us anything about whether Posner’s authorship score is significantly different from the next-ranked judge (Clay). When we compare Posner against Clay, we do not find any statistically significant difference. Nevertheless, assuming that authorship is a positive, part of our goal in ranking judges is to incen-
tivize all judges to exert greater effort in authoring their judicial opinions. Even if no statistically significant difference exists between any two particular judges, the ranking will induce judges to exert effort. So long as greater effort increases the likelihood that a judge will rank higher than another judge (even if not with statistical significance), the judge will have an incentive to exert more effort.92

Second, we examine invocation rates. An invocation is a citation whereby the judge is mentioned in the citing opinion by name (other than a perfunctory use of the name as, say, part of a parenthetical indication that the judge authored a particular dissenting or concurring opinion).93 Higher invocation rates for a judge’s opinions, we posit, correlate with a higher likelihood that the invoked judge authored her own opinions.

How are invocations related to authorship? We assume that judges have institutional knowledge as to which of their colleagues write their own opinions. Because an invocation represents a special indication of respect to the judge being cited (ordinarily judges are not referred to by name), it is unlikely that the special respect will be given unless the judge in question is one who writes her own opinions.94 Put differently, a judge who is known to delegate the majority of her opinions to the clerks

<table>
<thead>
<tr>
<th>RANK</th>
<th>JUDGE</th>
<th>CIRCUIT</th>
<th>AVERAGE SELF-CITATIONS TO EACH OPINION PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wollman, Roger L.</td>
<td>8</td>
<td>0.3393</td>
</tr>
<tr>
<td>2</td>
<td>Posner, Richard A.</td>
<td>7</td>
<td>0.3283</td>
</tr>
<tr>
<td>3</td>
<td>Bruce M. Selya</td>
<td>1</td>
<td>0.3230</td>
</tr>
<tr>
<td>4</td>
<td>Easterbrook, Frank H.</td>
<td>7</td>
<td>0.2520</td>
</tr>
<tr>
<td>5</td>
<td>Kelly, Paul J., Jr.</td>
<td>10</td>
<td>0.1873</td>
</tr>
<tr>
<td>6</td>
<td>Lynch, Sandra L.</td>
<td>1</td>
<td>0.1673</td>
</tr>
<tr>
<td>7</td>
<td>Carnes, Ed</td>
<td>11</td>
<td>0.1612</td>
</tr>
<tr>
<td>8</td>
<td>Moore, Karen Nelson</td>
<td>6</td>
<td>0.1600</td>
</tr>
<tr>
<td>9</td>
<td>Clay, Eric L.</td>
<td>6</td>
<td>0.1443</td>
</tr>
<tr>
<td>10</td>
<td>Coffey, John L.</td>
<td>7</td>
<td>0.1378</td>
</tr>
<tr>
<td>11</td>
<td>Garland, Merrick B.</td>
<td>DC</td>
<td>0.1343</td>
</tr>
<tr>
<td>12</td>
<td>Tacha, Deanell Reece</td>
<td>10</td>
<td>0.1334</td>
</tr>
<tr>
<td>13</td>
<td>Kanne, Michael S.</td>
<td>7</td>
<td>0.1331</td>
</tr>
<tr>
<td>14</td>
<td>Ebel, David M.</td>
<td>10</td>
<td>0.1296</td>
</tr>
<tr>
<td>15</td>
<td>Ripple, Kenneth F.</td>
<td>7</td>
<td>0.1246</td>
</tr>
<tr>
<td>20</td>
<td>Boudin, Michael</td>
<td>1</td>
<td>0.1016</td>
</tr>
</tbody>
</table>

Note that Posner and Easterbrook do even better than in the self-citation rate per opinion measure. Boudin also does better, but still remains outside the top fifteen.

92. Judges may write opinions of different length. The longer the opinion, the greater the likelihood that the opinion will receive a self-citation (due, for example, to the greater amount of analysis in a longer opinion). To control for this possibility, we also ranked the judges in our sample based on the average number of self-citations to each written opinion page. We provide the ranking below.

93. See Choi & Gulati, supra note 5, at 58-61 (defining invocations and analyzing judges based on the rates at which their names are invoked in other opinions).

94. See Landes et al., supra note 14, at 274 (“It is also not implausible that judges who write their own opinions will be more influential, since their opinions will be more consistent and, if good, then more consistently good than opinions written by law clerks.”).
is unlikely to find her judicial colleagues invoking her. They may cite the judge, if her opinion is on point, but will be unlikely to invoke her and thereby give her a special measure of respect.95

Invocation rates are at least somewhat robust to the subject matter of opinions. A judge who is held in high regard among other judges will tend to receive invocations for all types of opinions. Nonetheless, subject matter will not be completely irrelevant to the extent that judges do tend to invoke other judges for specific types of opinions (for example, Easterbrook on corporate and commercial-related law). Table 4 reports the results from the invocation rate ranking. A judge’s invocation rate is defined as the average number of invocations (measured from 1998 to 2003) of each judge’s opinions written in the 1998 to 2000 time period.

TABLE 4
RANKING BASED ON INVOCATION RATE

<table>
<thead>
<tr>
<th>RANK</th>
<th>JUDGE</th>
<th>CIRCUIT</th>
<th>AVERAGE INVOCATIONS TO EACH OPINION</th>
<th>T-TEST OF DIFFERENCE WITH THE MEDIAN JUDGE* (EQUAL VARIANCES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Posner, Richard A.</td>
<td>7</td>
<td>0.664</td>
<td>6.34**</td>
</tr>
<tr>
<td>2</td>
<td>Easterbrook, Frank H.</td>
<td>7</td>
<td>0.442</td>
<td>3.78**</td>
</tr>
<tr>
<td>3</td>
<td>Calabresi, Guido</td>
<td>2</td>
<td>0.228</td>
<td>3.08**</td>
</tr>
<tr>
<td>4</td>
<td>Wilkinson, J. Harvie, III</td>
<td>4</td>
<td>0.185</td>
<td>2.59**</td>
</tr>
<tr>
<td>5</td>
<td>Edmondson, J.L.</td>
<td>11</td>
<td>0.138</td>
<td>1.17</td>
</tr>
<tr>
<td>6</td>
<td>Higginbotham, Patrick E.</td>
<td>5</td>
<td>0.124</td>
<td>1.34</td>
</tr>
<tr>
<td>7</td>
<td>Luttig, J. Michael</td>
<td>4</td>
<td>0.124</td>
<td>1.14</td>
</tr>
<tr>
<td>8</td>
<td>Jones, Edith H.</td>
<td>5</td>
<td>0.109</td>
<td>0.98</td>
</tr>
<tr>
<td>9</td>
<td>Boudin, Michael</td>
<td>1</td>
<td>0.096</td>
<td>1.30</td>
</tr>
<tr>
<td>10</td>
<td>Walker, John M., Jr.</td>
<td>2</td>
<td>0.095</td>
<td>1.11</td>
</tr>
<tr>
<td>11</td>
<td>Clay, Eric L.</td>
<td>6</td>
<td>0.086</td>
<td>0.83</td>
</tr>
<tr>
<td>12</td>
<td>Cabranes, José A.</td>
<td>2</td>
<td>0.086</td>
<td>0.92</td>
</tr>
<tr>
<td>13</td>
<td>Kleinfeld, Andrew J.</td>
<td>9</td>
<td>0.085</td>
<td>0.86</td>
</tr>
<tr>
<td>14</td>
<td>Tjoflat, Gerald Bard</td>
<td>11</td>
<td>0.083</td>
<td>0.55</td>
</tr>
<tr>
<td>15</td>
<td>King, Carolyn Dineen</td>
<td>5</td>
<td>0.082</td>
<td>0.77</td>
</tr>
</tbody>
</table>

* The median judge (Terence T. Evans) is chosen as the forty-ninth ranked judge out of ninety-eight total judges.
** Indicates a significance level of 5%. The invocation rate for Terence T. Evans was 0.039 to each opinion.

95. On this special measure of respect that accrues to those judges doing their own work, Justice Brandeis famously said that “[t]he reason the public thinks so much of the Justices of the Supreme Court is that they are almost the only people in Washington who do their own work.” DAVID M. O’BRIEN, STORM CENTER: THE SUPREME COURT IN AMERICAN POLITICS 116 (5th ed. 2000) (internal quotation marks omitted); see also John G. Kester, The Law Clerk Explosion, LITIGATION, Spring 1983, at 20, 62.
All three of the test judges score in the top ten in terms of invocation rates. Indeed, Posner and Easterbrook are numbers one and two, respectively, in the ranking. Both Posner and Easterbrook’s scores, in addition, are significantly different from the median judge’s (Terence T. Evans) score. Posner’s and Easterbrook’s high t-statistics indicate that they are relatively more likely to self-author compared to the median judge in a statistically significant manner. Some evidence exists that a ranking based on invocation rates may distinguish among judges based on their self-authorship of judicial opinions. Nonetheless, reputation may reflect a judge’s long-term self-authorship pattern rather than the degree of authorship in any particular set of contemporary opinions.

Judges who write their own opinions, in addition to being more likely to cite themselves, might also be more likely to cite a smaller number of other judges relative to the average citation pattern. Judges are likely to have a more fine-tuned sense of which of the other judges are worthy of citation than their law clerks. Once a judge identifies her set of preferred other judges, the judge is likely to stick with them in her citation pattern, leading to a relatively low level of variance in citations. On the other hand, where clerks author the opinions, they will not necessarily cite to the same set of judges but may cite to widely differing judges. The variation in the number of different judges cited, therefore, will likely be higher with a judge who delegates extensively compared with a judge who tends to author her own opinions. Because our 1998 to 2000 dataset does not contain information on the specific identities of judges that a particular judge cites, we are unable to test whether variance in citation performs well in distinguishing our a priori set of judges who are reputed to self-author their opinions. We leave this test to another paper.96

2. Subject Matter-Neutral Language Patterns

The writing of judges who author their own opinions will likely display patterns. While specific word patterns (for example, use of the phrase “habeas corpus”) may be subject matter-specific, some patterns may not depend on the subject matter of a judicial opinion. The use of these genre-neutral patterns provides an alternative “white-box” method of testing for authorship.

For example, some judges will tend to write short opinions and others will write longer ones. Some will use extended quotes and others will not. When judges delegate to clerks, however, there will likely be greater variation in the types of opinions because clerks will

96. We are in the process of collecting these data now.
have their own styles. For some fixed number of opinions, a judge who writes her own opinions is likely to have a smaller opinion-size variation than a judge who does not. In our opinion, neither the length of the opinion (or average paragraph or sentence) nor the use of block quotes are necessarily tied to any specific subject matter or area of the law.

For each of the ninety-eight judges in our sample, we calculated the standard deviation of the length of each majority opinion based on published pages in the West Federal Reporter and excluding the summary and West keynotes. The standard deviation of majority-opinion length is for opinions from the 1998 to 2000 time period for each judge. Table 5 reports the judges’ rankings based on the standard deviation of majority-opinion length.

### Table 5

**Ranking Based on Standard Deviation of Majority-Opinion Length**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Judge</th>
<th>Circuit</th>
<th>Standard Deviation of Majority-Opinion Length</th>
<th>F-test P-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Loken, James B.</td>
<td>8</td>
<td>1.81</td>
<td>0.0000</td>
</tr>
<tr>
<td>2</td>
<td>Posner, Richard A.</td>
<td>7</td>
<td>1.84</td>
<td>0.0000</td>
</tr>
<tr>
<td>3</td>
<td>Martin, Boyce F., Jr.</td>
<td>6</td>
<td>1.84</td>
<td>0.0087</td>
</tr>
<tr>
<td>4</td>
<td>Wollman, Roger L.</td>
<td>8</td>
<td>1.92</td>
<td>0.0002</td>
</tr>
<tr>
<td>5</td>
<td>Easterbrook, Frank H.</td>
<td>7</td>
<td>1.99</td>
<td>0.0002</td>
</tr>
<tr>
<td>6</td>
<td>Black, Susan H.</td>
<td>11</td>
<td>2.21</td>
<td>0.0277</td>
</tr>
<tr>
<td>7</td>
<td>Arnold, Morris S.</td>
<td>8</td>
<td>2.25</td>
<td>0.0044</td>
</tr>
<tr>
<td>8</td>
<td>Boudin, Michael</td>
<td>1</td>
<td>2.33</td>
<td>0.0091</td>
</tr>
<tr>
<td>9</td>
<td>Ginsburg, Douglas H.</td>
<td>DC</td>
<td>2.37</td>
<td>0.0267</td>
</tr>
<tr>
<td>10</td>
<td>Hawkins, Michael Daly</td>
<td>9</td>
<td>2.51</td>
<td>0.0636</td>
</tr>
<tr>
<td>11</td>
<td>Schroeder, Mary M.</td>
<td>9</td>
<td>2.53</td>
<td>0.0586</td>
</tr>
<tr>
<td>12</td>
<td>Tashima, A. Wallace</td>
<td>9</td>
<td>2.64</td>
<td>0.0707</td>
</tr>
<tr>
<td>13</td>
<td>Walker, John M., Jr.</td>
<td>2</td>
<td>2.64</td>
<td>0.0638</td>
</tr>
<tr>
<td>14</td>
<td>Nygaard, Richard L.</td>
<td>3</td>
<td>2.67</td>
<td>0.1243</td>
</tr>
<tr>
<td>15</td>
<td>Sentelle, David B.</td>
<td>DC</td>
<td>2.73</td>
<td>0.1146</td>
</tr>
</tbody>
</table>

* The F-test provides a test of the null hypothesis that the standard deviation of opinion length for a particular judge is the same as the standard deviation for the median judge (Deanell Reece Tacha). The standard deviation of opinion length for Deanell Reece Tacha was 3.65.
As with the self-citation rate and invocation rate tests, Posner and Easterbrook score in the top-ten judges. In addition, Boudin—who was in the top ten for the invocation rate test but not for the self-citation rate test—is also ranked among the top-ten judges based on a lower standard deviation of opinion length. The differences between the standard deviation scores for Posner, Easterbrook, and Boudin compared with the score for the median judge (Deanell Reece Tacha) are significant at the one percent level. The standard deviation of the majority-opinion length for the judges successfully distinguished our test judges in terms of both high rank among our sample judges and comparison with the median judge.

We also hypothesize that judges who write their own opinions will produce shorter opinions with fewer quotes. They have a limited amount of time, and a large amount of work, and they are more confident about what they are saying. In Table 6, infra, we provide a ranking of our ninety-eight judges according to average length of majority opinions during the 1998-2000 time period. The length of each opinion is based on published pages in the West Federal Reporter excluding the summary and West keynotes. The average printed pages per majority opinion is equal to the total length of all opinions for a particular judge divided by the number of opinions written from 1998 to 2000.
### Table 6

**RANKING BASED ON AVERAGE PAGES PER MAJORITY OPINION**

<table>
<thead>
<tr>
<th>RANK</th>
<th>JUDGE</th>
<th>CIRCUIT</th>
<th>AVERAGE PRINTED PAGES PER OPINION</th>
<th>T-TEST OF DIFFERENCE WITH THE MEDIAN JUDGE* (UNEQUAL VARIANCES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Arnold, Morris S.</td>
<td>8</td>
<td>3.94</td>
<td>-10.14**</td>
</tr>
<tr>
<td>2</td>
<td>Easterbrook, Frank H.</td>
<td>7</td>
<td>4.38</td>
<td>-9.47**</td>
</tr>
<tr>
<td>3</td>
<td>Posner, Richard A.</td>
<td>7</td>
<td>4.44</td>
<td>-9.51**</td>
</tr>
<tr>
<td>4</td>
<td>Loken, James B.</td>
<td>8</td>
<td>4.48</td>
<td>-9.06**</td>
</tr>
<tr>
<td>5</td>
<td>Ginsburg, Douglas H.</td>
<td>DC</td>
<td>4.65</td>
<td>-7.22**</td>
</tr>
<tr>
<td>6</td>
<td>Wollman, Roger L.</td>
<td>8</td>
<td>4.66</td>
<td>-8.48**</td>
</tr>
<tr>
<td>7</td>
<td>Martin, Boyce F., Jr.</td>
<td>6</td>
<td>4.73</td>
<td>-6.84**</td>
</tr>
<tr>
<td>8</td>
<td>Schroeder, Mary M.</td>
<td>9</td>
<td>4.93</td>
<td>-6.20**</td>
</tr>
<tr>
<td>9</td>
<td>Black, Susan H.</td>
<td>11</td>
<td>5.04</td>
<td>-5.86**</td>
</tr>
<tr>
<td>10</td>
<td>Bowman, Pasco M.</td>
<td>8</td>
<td>5.17</td>
<td>-5.91**</td>
</tr>
<tr>
<td>11</td>
<td>Murphy, Diana E.</td>
<td>8</td>
<td>5.32</td>
<td>-5.27**</td>
</tr>
<tr>
<td>12</td>
<td>Evans, Terence T.</td>
<td>7</td>
<td>5.36</td>
<td>-5.76**</td>
</tr>
<tr>
<td>13</td>
<td>Higginbotham, Patrick E.</td>
<td>5</td>
<td>5.65</td>
<td>-4.49**</td>
</tr>
<tr>
<td>14</td>
<td>Kozinski, Alex</td>
<td>9</td>
<td>5.70</td>
<td>-3.17**</td>
</tr>
<tr>
<td>15</td>
<td>Edmondson, J.L.</td>
<td>11</td>
<td>6.00</td>
<td>-3.08**</td>
</tr>
<tr>
<td>16</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>17</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>18</td>
<td>Boudin, Michael</td>
<td>1</td>
<td>6.09</td>
<td>-4.17**</td>
</tr>
</tbody>
</table>

* The median judge (Paul V. Niemeyer) is chosen as the forty-ninth ranked judge out of ninety-eight total judges. Niemeyer had an average majority opinion length of 7.67 pages.

** Indicates a significance level of 5%.

Posner and Easterbrook again appear among the top-ten judges in terms of likelihood of self-authorship. Boudin, however, falls out of the top ten and is ranked number eighteen, but still within the top-twenty judges. The average majority-opinion page length for all three test judges was significantly lower than the average page length for the median judge (Paul V. Niemeyer) as deduced using a two-sided t-test assuming unequal variances.

Other frequency-based tests may prove feasible. A judge writing an opinion will be less likely to footnote her opinion heavily than the law clerk. 97 Clerks, because of their lower knowledge base, higher level of insecurity, and law review training, are more likely to feel a need to footnote the document. We predict, therefore, that a lower

---

97. See supra note 70 and accompanying text (citing Posner, who hypothesizes that clerks will use more footnotes).
number of footnotes per opinion (or per page) correlates with a higher likelihood of self-authorship.98

3. Revisiting the Black-Box Tests

We and others may learn from the failure of our initial run at the “black-box” GZip-based tests. In order to take advantage of genre-specific information and to control for genre-specific commonalities across opinions written by different authors, at least three changes to our methodology are possible.

First, to control for the possibility that a particular clerk may dominate a judge’s opinion writing for a specific year, the sample of opinions for each judge should span a relatively long time period—certainly greater than the three years covered by our 1998 to 2000 time period. Increasing the number of opinions for each judge—from four to ten, for example—may also provide traction to the GZip results.

Second, as discussed above, we were too zealous in our desire to remove unnecessary information from our opinions. The informativeness of the opinions (on the issue of authorship) would likely have been greater if we had kept the citation information. Keeping citation information also greatly reduces the work required in preparing opinions for the GZip-based tests.

Third, a study could control for the noise generated from comparing opinions across different areas of the law. Suppose Judge A writes her own opinions while Judge B does not. If we compared four

98. We used two other less sophisticated methods based on the publicly available Litstat program. For the Litstat tool on the Internet, see Matthew Bielich, LitStat, at http://www.pinionsolutions.com/litstat/ (last visited Jan. 22, 2005). Authors with distinctive styles and patterns are assumed to use similar types of sentence and word structures in all of their writing. Using the Litstat program, we calculated the alphabetical frequency of a text, that is, the frequency with which the words in that text begin with a certain letter of the alphabet. Frequency of commonly used words is an indicator of the specificity of style. Thus, each author is likely to have a set of words that he or she may use more frequently than other authors. A frequency distribution of words could thus be an indicator of the stylistic uniqueness. For each judge we used two samples of text of 8000 characters each. As with our other black-box tests, we did not control for the subject matter of the judicial opinions. As with the GZip test, we cleaned the text. Here, however, we did not take out common words such as “and” and “the,” because the information as to those commonly used words is actually important to these supplementary tests. The alphabetical distribution of the words in the opinions is the output of the tool. This output is copied to a spreadsheet, and the average of the difference of these values for the two opinions is calculated—the lower the average, the higher the probability of sole authorship. None of our three control judges were in the top fifteen judges in terms of probability of sole authorship. We also used the Litstat program to calculate the average sentence length in a fixed amount of text (8000 characters, after the text has been cleaned for citations, quotes, West headnotes and other extraneous information). We predicted that the variance in these frequencies would be smaller among documents authored by the same author than among those authored by different authors. Only Boudin scored in the top fifteen judges (at number two) in terms of probability of sole authorship.
opinions by Judge A on diverse areas of the law (including, for example, criminal law, bankruptcy law, administrative law, and securities regulation) against four opinions by Judge B on securities regulation, Judge B will likely receive a higher same-authorship score. Such a pattern, possible where opinions are randomly selected without regard to subject matter, may introduce noise into the analysis.

A simple control for such noise would be to select randomly opinions from the same subject matter area of the law—for example. It is possible that the common phrases in criminal law opinions (for example, “habeas corpus”) are so frequent that all judges will receive the identical same-authorship score even with this control. Nonetheless, to the extent that, after taking into account the common baseline vocabulary, judges do have idiosyncratic writing styles, the GZip methodologies should pick up on these differences in determining authorship.

We leave implementing these white-box controls to the black-box tests for the genre-specific nature of opinions for our next paper.

IV. CONCLUSION: OTHER POSSIBLE APPLICATIONS OF AUTHORSHIP TECHNOLOGY

This Essay only touches the surface of the world of technology that might be applied to determine judicial authorship of opinions. The results of the white-box tests are but a preliminary step toward determining authorship. Even assuming that these white-box tests have a significant amount of explanatory power to determine authorship, each of the white-box tests measures something that, at best, correlates with authorship. We suspect that meaningful progress in this area will require a serious collaborative research effort between legal academics and scholars in computational linguistics. Such technology, once developed, has the potential to assist in areas other than research on the judiciary. We list a couple below.

A. Securities Fraud Complaints

Congress enacted the Private Securities Litigation Reform Act of 1995 to combat frivolous litigation. Proponents of the Act argued that plaintiffs’ attorneys literally cut-and-pasted complaints together to file suit against any company that experienced a large drop in stock price, regardless of whether any real evidence of fraud ex-
listed.\textsuperscript{100} Congress intended the Reform Act to force plaintiffs’ attorneys to take more care in their filings and provide particularized factual assertions as to why fraud in fact exists.

We may be able to use the authorship methodology to determine how “close” two securities fraud complaints are to one another. If a complaint is really a cut-and-paste job, the complaint is likely to receive a high same-authorship score when compared to other securities complaints from the same law firm. If the Reform Act resulted in more particularized investigation before the filing of a complaint, we predict that a set of pre-Reform complaints from the same law firm would produce a higher same-authorship score as compared to a set of post-Reform complaints from the same law firm.

\textbf{B. Boilerplate Contract Evolution}

The contracts used in a variety of settings such as the corporate bond area are commonly described as boilerplate. In other words, essentially the same standard language gets repeated in every contract in the market. Even if individual drafters do not know what the contract language means, it gets repeated because everyone else in the market is using it. Problems can arise when, in the course of a dispute, the two sides assert different meanings for some piece of ambiguous language whose original understanding has long since been lost.

Authorship testing programs may be used to trace the original source of the later contracts. Finding the source is potentially important for at least two reasons. First, from a practical dispute-resolution point of view, knowing the original understanding might help resolve the current dispute (to the extent the judge decides that the original understanding should govern). Second, from a research point of view, the ability to track the evolution of contract language can help us understand how contract language evolves, how it becomes boilerplate, and what circumstances produce changes in boilerplate.\textsuperscript{101}


EMPIRICAL MEASURES OF JUDICIAL PERFORMANCE: THOUGHTS ON CHOI AND GULATI’S TOURNAMENT OF JUDGES

BRANNON P. DENNING*

I. EXAMINING THE ASSUMPTIONS IN A TOURNAMENT OF JUDGES? ........................................ 1125
   A. Assumption No. 1: Politics Should Play a Very Small Role in the Selection of Supreme Court Nominees .................................................. 1125
   B. Assumption No. 2: The Present Level of Politicization in the Supreme Court Appointments Process is Aberrant ................................................................. 1128
   C. Assumption No. 3: Claims of “Merit” by Presidents for Their Court Nominees Obfuscate the Politicization of the Process and Mislead the Public ........................................ 1130
   D. Assumption No. 4: Tournament Would Introduce Needed Transparency to Selection Process and Assumption No. 5: Possibility of Ex Ante Agreement on Objective Criteria ........................................................................ 1133

II. HOW TOURNAMENTS CAN AID SUPREME COURT SELECTION ........................................ 1134
   A. Improving the President’s Selection Process ...................................................................... 1135
   B. Creating a Senate “Advice” Mechanism ............................................................................. 1137
   C. Ending the ABA’s Monopoly on Candidate Vetting .......................................................... 1139
   D. Creating a New Vocabulary for Debating Nominations ...................................................... 1140

III. CONCLUSION .................................................................................. 1142

Stephen Choi and Mitu Gulati have constructed a fascinating thought experiment1 to restore rationality to a process that was once referred to as the confirmation “mess”2 but that is now described as “war.”3 They have, moreover, put their tournament into operation and composed a list of court of appeals “winners” from whom future Supreme Court nominees might be chosen and for whom the President could make credible claims of selection based on merit.4

As someone who has written about the pathologies of the confirmation process5 and has suggested reforms that are not particularly

---

* Associate Professor of Law, Cumberland School of Law at Samford University. I thank Mitu Gulati for sharing his and Stephen Choi’s excellent work with me and for suggesting me for this Symposium. Thanks also to Jim Rossi for the kind invitation to contribute, and to Bill Ross for reading and commenting on early drafts.

viable politically. I can sympathize with Choi and Gulati’s sense that “the present Supreme Court selection system is so abysmal that even choice by lottery might be more productive” as well as their desire to find a solution, even one unlikely to be adopted in the form they propose. However, the tournament they propose is based on a number of unelaborated assumptions about the state of the Supreme Court appointment process that bear closer scrutiny. Despite my skepticism about some of their assumptions, however, I think that a form of tournament could be extremely useful to all stakeholders in the confirmation process—the President, the Senate, interest groups, and the public.

Part I of this Essay examines the assumptions driving Choi and Gulati’s proposed tournament of judges. I conclude that those assumptions may not be correct, or at least that they require some elaboration to support the strong claims that Choi and Gulati make. My criticisms are intended not to dismiss the proposal, but rather to encourage Choi and Gulati to refine it. They have already shown that a tournament can be run fairly easily, and that it can produce some surprising results. Explaining and defending their assumptions may increase the possibility that their proposal is taken up by participants in the process. In Part II, I argue that if we abandon an extreme form of the tournament, that is, “one that bars the president and the Senate from putting forth merit-related rationales outside [a] list of objective factors,” a tournament may be extremely useful to many of the interested parties in the selection and confirmation process.

6. See, e.g., Denning, Reforming the New Confirmation Process, supra note 5, at 31-41 (proposing reforms to the confirmation process, including amending Senate filibuster rules, curbing power of committee chairs, curbing the use of the “hold,” and constructing a meaningful “advice” mechanism for Senators); Denning, Blue Slip, supra note 5, at 97-101 (proposing reforms to the “blue slip” procedure in the Senate Judiciary Committee, including making it a formal part of the Committee’s rules).
7. Choi & Gulati, supra note 1, at 301.
8. Id. at 322 (“How likely is it that a proposal like ours would get implemented? In the form we suggest, there is no chance.”).
9. Please note that the brevity of the original essay is a point in its favor, and the authors were quite explicit about wanting to open a debate. Choi and Gulati noted: Our essay serves only as a starting point. We hope the potential benefits of exposing the politics involved in the selection of Supreme Court justices to greater scrutiny as well as introducing greater competition among appellate court judges will lead to an ongoing debate about how to reform our judicial system.
10. See Choi & Gulati, supra note 4.
11. Id. at 80-81 (describing the surprisingly high performance of ranking judges reputed to be on Bush’s short list for Supreme Court nominations).
12. Choi & Gulati, supra note 1, at 313.
process. I doubt, though, that it could be as transformative as Choi and Gulati sometimes suggest. A brief conclusion follows in Part III.

I. EXAMINING THE ASSUMPTIONS IN A TOURNAMENT OF JUDGES?

A close reading of Choi and Gulati’s initial essay reveals five assumptions driving their proposal: (1) politics should play a very small role in the selection of Supreme Court nominees; (2) the present level of politicization in the Supreme Court appointments process is aberrant and produces bad selections; (3) claims of “merit” by Presidents for their Court nominees obscure the politicization of the process and mislead the public; (4) a tournament will bring transparency to the process, exposing and diminishing the degree of politicization; and (5) nominees’ opponents and proponents can agree ex ante on objective criteria for the tournament. I argue in this Part that each of these assumptions either is questionable as a descriptive matter or requires a stronger normative defense than Choi and Gulati offer.

A. Assumption No. 1: Politics Should Play a Very Small Role in the Selection of Supreme Court Nominees

Choi and Gulati repeatedly contrast “intellectual ability,” or “merit,” with “politics” and regard the two as virtually mutually exclusive selection criteria. One aim of their tournament is to reduce the role that politics plays in nominations to the Supreme Court. From this starting point, three additional questions arise, which are not answered by Choi and Gulati. The first is definitional: what precisely do they mean by “politics”? Second, why exactly should we seek to eliminate politics, however defined, from the process? Third, if, as they concede, total elimination of politics from the process is unlikely, how much politics is optimal or acceptable?

Portions of their essay suggest that Choi and Gulati equate “politics” with “ideology” in the sense that selections are now made according to how candidates are expected to vote on a set of discrete issues. “The current selection criteria for the Supreme Court,” they

13. See generally id.
15. See, e.g., id. at 303 (arguing that selecting “future Supreme Court justices on the basis of . . . objective criteria would make clear (and thereby reduce) the role that politics plays in both the initial process of selecting a candidate and the often highly political Senate confirmation proceedings”).
16. See, e.g., id. at 302 (“Our best guess is that politicians define ‘merit’ in terms of ideology, and argue accordingly. We suggest that a market-based system would be an improvement.”). In a subsequent essay, they clarify this a bit by contrasting “merit,” which “refers to more widely held views of what makes a good judge,” with “ideology,” which “refers to narrowly held views.” Choi & Gulati, supra note 4, at 27. As Choi and Gulati acknowledge, it is still difficult to define ideology. Id. at 27 n.6. Moreover, the substitution of
write, “appear to be a set of political litmus tests on matters such as abortion, the death penalty, and affirmative action.”17 But then why not use “ideology”? After all, describing the selection process as “political” could mean other things—including (1) nominations calculated to appeal to certain interest groups or to members of the Senate and (2) nominations to repay presidential debts to political allies—or it could describe campaigning by or on behalf of prospective nominees. Indeed, elsewhere Choi and Gulati seem to condemn these manifestations of politics too.18

Lest I be seen as nitpicking, I think that a precise definition of the “politics” that Choi and Gulati condemn is necessary because of the normative question lurking in the background: Why should the confirmation process be insulated from politics? They seem to identify two problems with an overtly political selection process. First, there is the harm to judicial independence—political actors will seek to pack the Court with ideological proxies who will do their nominator’s bidding.19 Second, there is, they argue, harm in courts being seen not as impartial adjudicators but as employing a neutral façade to justify results predetermined by the deciding Justice’s ideological commitments.20

But there is another way of understanding “political” judicial selection. The two main players—the President and the Senate—are popularly elected and therefore presumably responsive to their constituents. When a President nominates someone to the U.S. Supreme Court or when the Senate is asked to confirm that nominee, each is really deciding whether to confer on a nominee unreviewable discretion to be exercised for as long as that person chooses to remain on the Court. The small number of cases that the Supreme Court now takes often places it squarely in the middle of highly contested cultural conflicts—abortion, affirmative action, sexual privacy—and it resolves them outside the legislative process. At this level the legal materials are the least determinant, and resolution of the legal questions often requires the Court to weigh conflicting constitutional values. If de Tocqueville was correct that most political questions in

17. Choi & Gulati, supra note 1, at 305 (emphasis added).
18. Id. at 312 (“Underlying most appointments, we suspect, are political agendas and the repayment of political favors.”). For similar observations to those made above, see Michael J. Gerhardt, Merit vs. Ideology, 26 CARDozo L. REV. 325, 327 (2005).
19. Choi & Gulati, supra note 1, at 300 (“An effective constitutional democracy requires an independent judiciary.”); id. at 310-11 (discussing the importance of judicial “independence,” defined as deciding cases “impartially,” that is, “independent of political ideology,” and arguing that “[e]vidence suggests that judges fall short of this mark”).
20. Id. at 317 (“[A] tournament system should help reduce the appearance of political bias that exists today. Evidence reveals that who you get as a judge, in terms of political affiliation, can often affect the outcome of the case.”).
America become legal ones, then the Court is as much a political institution as the Presidency or Congress.\textsuperscript{21} Is it so wrong, then, for the President or Senate to seek some assurances—for instance, that the nominee thinks like the President or that a nominee is not outside the mainstream and might tip the balance on a closely divided Court—before judicial nominees are elevated to the bench and placed outside the ordinary electoral accountability? If this is the politics that Choi and Gulati want to eliminate from the selection process, they need to provide a more compelling normative justification.

Choi and Gulati seem to recognize that eliminating all politics, however defined, from the selection process is not possible. Presidents, to a greater or lesser extent, have sought out nominees with whom they felt they shared certain values.\textsuperscript{22} The Senate has often invoked “political”—even ideological—concerns when rejecting certain High Court nominees.\textsuperscript{23} And yet the Supreme Court has not only survived, it remains one of the most revered organs of our government, even after \textit{Bush v. Gore}.\textsuperscript{24}

Choi and Gulati admit that their proposal “will not eliminate political considerations;”\textsuperscript{25} they even write that they are “not necessar-

\textsuperscript{21} It is important to remember that the subject of Choi and Gulati's essay is selection for the \textit{Supreme Court}, despite their article's use of the word "judge" throughout. This is not to say that appeals court judges do not have similar discretion, but I would argue that they are at least not at liberty to overrule or disregard binding Supreme Court precedent, though they may have considerable discretion implementing Court decisions that are vague or ambiguous. For example, two recent decisions from the Eleventh Circuit Court of Appeals upholding Florida's statutory prohibition on homosexual adoption and Alabama's prohibition on the sale of sex toys, both decided after \textit{Lawrence v. Texas}, 539 U.S. 558 (2003), show how much discretion lower courts have when implementing Supreme Court decisions. See \textit{Williams v. Attorney Gen.}, 378 F.3d 1232 (11th Cir. 2004), \textit{cert. denied}, 125 S. Ct. 1335 (2005); \textit{Lofton v. Sec'y of the Dept of Children & Family Servs.}, 358 F.3d 804 (11th Cir. 2004), \textit{cert. denied}, 125 S. Ct. 869 (2005). The same phenomenon can be seen in many lower courts' treatments of the Court's recent Commerce Clause decisions. See, e.g., Brannon P. Denning & Glenn H. Reynolds, \textit{Rulings and Resistance: The New Commerce Clause Jurisprudence Encounters the Lower Courts}, 55 ARK. L. REV. 1253 (2003); Glenn H. Reynolds & Brannon P. Denning, \textit{Lower Court Readings of Lopez, or What if the Supreme Court Held a Constitutional Revolution and Nobody Came?}, 2000 WIS. L. REV. 369.


\textsuperscript{23} Abraham, \textit{supra} note 22, at 28-34 (discussing Senate rejection of nominees); Carter, \textit{supra} note 2, at 62-65 (discussing opposition to Thurgood Marshall's nomination); Mark Silverstein, \textit{Judicious Choices: The New Politics of Supreme Court Confirmations} 10-32 (1994) (discussing opposition to Abe Fortas's elevation to Chief Justice).

\textsuperscript{24} See, e.g., James L. Gibson et al., \textit{The Supreme Court and the US Presidential Election of 2000: Wounds, Self-Inflicted or Otherwise?}, 33 BRIT. J. POL. SCI. 535, 541 (2003) (arguing that a "reservoir of good will" built up by the Court enabled it to weather criticism over its resolution of the 2000 presidential election); see also \textit{Bush v. Gore}, 531 U.S. 98 (2000).

\textsuperscript{25} Choi & Gulati, \textit{supra} note 1, at 312.
ily against the inclusion of politics per se in the judicial nomination process but merely want trade-offs (which they assume to exist) between politics and merit to be transparent. But these concessions raise a denominator problem: How much politics is too much, or what is the optimal amount the process should tolerate?

B. Assumption No. 2: The Present Level of Politicization in the Supreme Court Appointments Process is Aberrant

Whatever the optimal amount that politics should play in the process, Choi and Gulati argue that politics is too prevalent in Supreme Court selection. They further suggest that current levels of politicization are aberrant and that a less politicized selection process would produce Justices who are better qualified than those now chosen.

The lack of an acceptable-level-of-politics baseline makes the claim that the current process is “overwhelmed by politics” difficult to assess. In his historical survey of Supreme Court selection and nomination, Henry Abraham wrote of Presidents’ consistent concerns with a “candidate’s real politics”—regardless of nominal party affiliation.

The chief executive’s crucial predictive judgment concerns itself with the nominee’s likely future voting pattern on the bench, based on his or her past stance and commitment on matters of public policy insofar as they are reliably discernible. All presidents have tried, thus, to pack the bench to a greater or lesser extent.

Even George Washington, in an age before the official emergence of parties, “not only had a septet of criteria for Court candidacy, but he adhered to them predictably and religiously,” wrote Abraham. David Yalof’s recent study of presidential Supreme Court selection shows every President from Truman through Clinton making politi-

26. Id. at 305.
27. Id. at 303-04, 305, 312-13, 316, 318, 321 (referring to transparency).
28. Id. at 304 (describing the current selection system as “a biased and nontransparent process overwhelmed by politics”).
29. Id. at 301 (calling the “present Supreme Court selection system . . . abysmal” and blaming “politics” for its sorry state); id. at 304 (defending tournament proposal by “point[ing] out how badly the present selection system works without [objective] measures”).
30. Id. at 304.
31. Abraham, supra note 22, at 49.
32. Id.
33. Id. at 53. They were: (1) support and advocacy of the Constitution; (2) distinguished service in the Revolution; (3) active participation in the political life of state or nation; (4) prior judicial experience on lower tribunals; (5) either a “favorable reputation with his fellows” or personal ties with Washington himself; (6) geographical suitability; (7) love of our country.

Id.
cical choices, either by instructing subordinates to find nominees whose “real politics” match the President’s (Eisenhower, to an extent; Nixon; and Reagan) or by nominating personal friends whom the President trusts (Truman, Kennedy, and Johnson). If later Presidents, like Nixon and Reagan, have become more concerned with positions that candidates are likely to take with regard to specific issues which come before the Court (like abortion), to the point of constructing disqualifying “litmus tests,” perhaps that is a rational response to the increased role the Court has played in resolving contentious social issues.

And yet, Abraham noted, “although a number of rather weak nominations have slipped past the Senate,” Presidents “have avoided nominating patently unqualified individuals to the high tribunal.” Of course that does not mean that Presidents could not have made even “better” selections, but it is impossible to know how much better those not chosen would have performed on the Court or to know how many “better” candidates were available to the President. That weak candidates are sometimes selected and confirmed does not go very far in establishing the truth of Choi and Gulati’s charge that politics has overwhelmed the selection process and that inferior selections are made as a result.

34. YALOF, supra note 22.
35. See Choi & Gulati, supra note 1, at 305 (“The current selection criteria for the Supreme Court appear to be a set of political litmus tests on matters such as abortion, the death penalty, and affirmative action.”). Democratic Presidents, too, can have litmus tests. In April 2004, Democratic presidential nominee John Kerry publicly announced that he would appoint only pro-choice judges to the Supreme Court. See Mike Glover, Kerry Underscores Stance on Abortion, CINCINNATI POST, Apr. 23, 2004, at A7. Subsequently, however, he softened that stance somewhat, suggesting that he might consider a pro-life judge, as long as Roe v. Wade is not threatened. See Jim Vandehei, Kerry Forced to Address Abortion, PITTSBURGH POST-GAZETTE, June 6, 2004, at A11 (noting concern among supporters at Kerry’s remarks).
36. ABRAHAM, supra note 22, at 33.
37. Learned Hand and Henry J. Friendly are the usual choices for great judges who never made it to the Supreme Court. See, e.g., id. (“What a pity that a man of Hand’s intellect and pen never became a member of the Supreme Court—he would have graced it from every point of view.”); id. at 214-15 (noting Lyndon B. Johnson’s refusal to nominate the “legal giant[]” Henry Friendly to the Supreme Court). Yalof’s study indicates that Presidents bypassed candidates with arguably better paper credentials for various reasons. See, e.g., YALOF, supra note 22, at 67 (noting that Eisenhower “excluded a large number of qualified candidates from consideration either because they were too young, held jobs in private practice, served on trial courts, or worked for the administration”); id. at 79 (identifying former Attorney General Robert Kennedy as “the main conduit for [President Kennedy] in the selection process” and noting that his “strenuous objections” to Paul Freund “buried Freund’s candidacy even while it enjoyed the support of at least four other high-level advisors”).
38. Abraham’s study notes that the Senate has not balked at rejecting nominees thought to be unqualified. Harold Carswell was the last such nominee, despite Senator Roman Hruska’s plea for the establishment of a “mediocre” seat on the Court. ABRAHAM, supra note 22, at 11 (describing Hruska’s “pathetic fumbling attempt to convert [Carswell’s] mediocrity into an asset”).
C. Assumption No. 3: Claims of “Merit” by Presidents for Their Court Nominees Obscure the Politicization of the Process and Mislead the Public

The prevalence of politics and the deleterious effects it has on the selection process, Choi and Gulati claim, are “disguised by claims about a particular candidate’s ‘merit.’”39 A tournament conducted by reference to ex ante criteria, they contend, will expose the gap between claims of merit made on behalf of nominees by their proponents and those who won the tournament. “It becomes hard to argue that candidate X is the ‘best’ candidate on purely merit grounds,” they write, “when preexisting objective criteria presumptively support candidate Y.”40 If candidate X is chosen, then the burden of explanation should fall on the President. “The public should know when a judge has been chosen because of ideology, and not because she measures up when objectively compared with contemporaries.”41

The argument for transparency itself entails a few questionable, or at least unverified, assumptions. First, assuming that “the public” is the audience for the information furnished by the tournament,42 the authors must presume a role for the public in selecting Supreme Court nominees. Officially, of course, the public plays no direct role in selecting or confirming Supreme Court Justices.43 Indirectly, however, Presidents who consistently make shoddy nominations, or Senators who confirm those nominations (or refuse to confirm worthy ones), could be held accountable by the voters.44 Moreover, political scientists have found that when deciding whether to support or oppose controversial Court nominees, Senators will endeavor to gauge

39. Choi & Gulati, supra note 1, at 301.
40. Id. at 304 (emphasis added).
41. Id. at 305. “Where political motivations drive the selection of an alternative candidate, our proposed system . . . would make it more likely that such motivations would be exposed to the public.” Id. at 304; see also id. at 305 (“[W]here politics does impact the selection of judges, such motivations should be made transparent to the public.”).
42. The authors are somewhat unclear on this point. While they frequently refer to the benefits accruing to the public from a transparent process, see, e.g., id. at 303, 305, 312, 313, they also refer to transparency encouraging dialogue between the President and the Senate, id. at 313, and even providing a means for journalists to evaluate candidates, id. at 322.
43. See U.S. CONST. art. II, § 2, cl. 2 (prescribing presidential nomination and Senate confirmation).
44. Indeed, this accountability was offered by Alexander Hamilton in The Federalist as a check on the potential for cronyism by nominating Presidents and partisan obstructionism by Senators. See THE FEDERALIST NO. 77, at 428-30 (Alexander Hamilton) (Clinton Rossiter ed., 1961). For evidence that votes against nominees made by popular Presidents can come back to haunt Senators, see Jeffrey A. Segal, Albert D. Cover & Charles M. Cameron, The Role of Ideology in Senate Confirmation of Supreme Court Justices, 77 KY. L.J. 485 (1988-89).
public support for the President, the nominee, or both.\footnote{45} Perhaps Choi and Gulati believe that for the public to form opinions on nominees in order to exercise that indirect influence, the public needs the comparative information that their tournament provides.

But perhaps they simply overestimate the degree to which voters and Senators credit presidential claims of merit. Political scientists James Gimpel and Lewis Ringel have noted that while public supporters of a particular nominee tend to tout the nominee’s objective qualifications—character, achievements, abilities, and the like—opponents tend to frame arguments against the nominee in terms of the nominee’s perceived ideology or expected positions on key issues.\footnote{46} In another paper, Gimpel argued that one of the strongest “cues” for those holding opinions about nominees to the Supreme Court is the nominating President.\footnote{47} “Public evaluations of the president’s choice for the Court,” they conclude, “are shaped by judgments of presidential performance. . . . The president serves as a cognitive link between the citizen and the Court in the judicial selection process.”\footnote{48} This link, in turn, can serve as a cue to Senators as to whether they can safely oppose a presidential nominee.\footnote{49} And when Senators are free to vote their preference, according to Segal, Cover, and Cameron, they tend to want to vote against nominees from whom they perceive themselves as ideologically distant and vice versa.\footnote{50}

\footnote{45. See, e.g., Kathleen Frankovic & Joyce Gelb, Public Opinion and the Thomas Nomination, 25 PS: POL. SCI. & POL. 481, 483 (1992); L. Martin Overby et al., Courting Constituents? An Analysis of the Senate Confirmation Vote on Justice Clarence Thomas, 86 AM. POL. SCI. REV. 997 (1992). Frankovic and Gelb wrote: By two to one, Americans thought Clarence Thomas should be confirmed before the hearings [on Anita Hill’s sexual harassment allegations]. After they were over, they still favored confirmation. . . . Many of the Senators decided to vote to confirm in the last hours before the vote. Multiple public opinion polls, done by media organizations, were available to Senators. Many surely used those polls to help make up their minds, or to justify minds they had already made up. The Senate was confused about who was telling the truth. Given the sensitivity of male Senators to a sexually charged confrontation, they sought to validate their position by recourse to the public’s judgment. Frankovic & Gelb, supra, at 483.}


\footnote{47. James G. Gimpel & Robin M. Wolpert, Opinion-Holding and Public Attitudes Toward Controversial Supreme Court Nominees, 49 POL. RES. Q. 163 (1996).}

\footnote{48. Id. at 173.}

\footnote{49. Id. at 174 (“It is therefore not surprising that senators feel free to give Supreme Court nominees a harder time when the appointing president is suffering at the polls.”).}

\footnote{50. Segal, Cover & Cameron, supra note 44, at 485; see also Lee Epstein et al., The Role of Qualifications in the Confirmation of Nominees to the U.S. Supreme Court, 32 FLA. ST. U. L. REV. 1145, 1168-72 (2005).}
In other words, if the political scientists are correct, Supreme Court nominees—all things being equal—will be confirmed if more Senators find the nominee closer to them ideologically than find the nominee ideologically distant. On the other hand, a President’s popularity may cause some Senators to cast votes counter to their ideological preference, lest they suffer the wrath of voters at a later date. The public, meanwhile, will tend to judge a nominee based on its approval or disapproval of the President.52 Supporters will tend to focus on general characteristics—many of which speak to a nominee’s merit—while opponents will tend to focus on specific issues and on ideology generally.52

Perhaps more interesting is evidence of the willingness of members of the public, when polled, to express their opposition to particular nominees in ideological terms—for example, “issue-specific concerns such as abortion and affirmative action”—as compared to evidence that members of the public who support particular nominees are more likely to state general reasons for their approval—for example, “the nominee’s personal traits and experience.”53 This suggests a certain sophistication on the part of engaged members of the public, who are aware that claims of merit for a particular nominee will be made in the larger context of a nominee’s ideological acceptability to the President. Certainly such an understanding exists among Senators, who often point out, when opposing a particular candidate, that they do not expect that the President (especially if the President is a member of another party) would nominate the same person they would.

None of these accounts, however, suggests that Presidents’ mere claims about a nominee’s merit are very persuasive, other than by virtue of the fact that the claims are being made by a popular President. Given the intense interest group activity that now surrounds Supreme Court nominations, members of the public for whom Supreme Court nominations are salient (and this still probably constitutes a very small percentage of “the public”) will be subjected to a barrage of information—both positive and negative—about a nominee other than a President’s statements in support. It does not seem plausible that these members of the public would be taken in or misled by presidential claims of merit.

Senators called upon to vote on the nomination will be subject to even more pressure and will have even more information to influence their vote. Even assuming that some members of the public might accept an administration’s claims of merit, Senators who wish to see

51. See Gimpel & Wolpert, supra note 47, at 171-73.
52. See Gimpel & Ringel, supra note 46, at 139-40.
53. See, e.g., id. at 139-41.
a particular nominee defeated can use the information at their disposal, as well as the national forum their office provides, to publicize negative aspects of a nominee’s record or question a nominee’s qualifications.54 Senator Ted Kennedy did this effective after the Bork nomination by making a speech on the floor of the Senate vehemently denouncing “Robert Bork’s America.”55

One final thought: If members of the public at large are formulating opinions on Supreme Court nominees that are rooted in ideology and if, as the evidence suggests, undecided members of the Senate take heed of public opinion when deciding whether to support a nominee, perhaps the “politics” in the selection process that Choi and Gulati decry56 comes as much from the bottom up as from the top down. That is, if members of the public understand there to be an ideological component to judging and to judicial selection and people know that courts (especially the Supreme Court) have a role in resolving issues important to them, then perhaps Presidents and Senators will be forced to take account of public expectations as never before. Thus, we see Democratic presidential candidates vowing to appoint only judges solicitous of Roe v. Wade57 and Republican Presidents pledging not to appoint “activist” judges.

D. Assumption No. 4: Tournament Would Introduce Needed Transparency to Selection Process and Assumption No. 5: Possibility of Ex Ante Agreement on Objective Criteria

Assuming that there is a level of confusion on the part of the public regarding Supreme Court selection that needs to be dispelled, would a tournament help? Choi and Gulati argue that a tournament would shed light on what they see as “a biased and nontransparent process.”58 But the success of a tournament in opening the selection process to scrutiny depends on the ability of potential antagonists in the process to agree, ex ante, on the “objective” criteria to which Choi and Gulati make repeated reference.59 If the selection process is already the subject of unhealthy politicization, as Choi and Gulati charge, one wonders what hope there is of getting Presidents, Senators, interest groups, journalists, academics, and members of the

54. Segal, Cover, and Cameron offer evidence that Senators may oppose the nominee of even a popular President if they can raise questions about qualifications. See Segal, Cover & Cameron, supra note 44, at 485.
56. See Choi & Gulati, supra note 1, at 301.
57. See supra note 55 and accompanying text.
58. Choi & Gulati, supra note 1, at 301. For references to “transparency,” see supra note 27.
59. Id. passim.
public to agree on what objective criteria ought to be used and how they should be applied to potential nominees.

A possible response is that the transparency inherent in a tournament, as Choi and Gulati suggest, permits criticism of the criteria, the criteria’s application, or both. However, to improve on an existing tournament, one need only use different criteria or apply the existing criteria differently. As improvements result in the formulation of alternate tournaments, such tournaments would themselves be transparent and open to criticism, which could lead to the fashioning of more tournaments, and so on. There could then conceivably be a Tournament among tournaments, with consumers of the information making choices among them. In that way, we might eventually reach a consensus regarding the set of objective criteria by which judges are to be measured.

But in the short run, this “let a thousand flowers bloom” approach would seriously reduce the possibility of dispelling confusion surrounding the selection process. Those lacking independent means for discriminating among the competing tournaments might simply resort to rough proxies. For example, liberal Presidents and Senators might use tournaments run by organizations such as the American Constitution Society or People for the American Way, while conservatives might turn to the Federalist Society, the CATO Institute, or the Heritage Foundation. Even if there was an overlap in the “objective” criteria employed by these different organizations, I strongly suspect that there would be little overlap among the lists of presumptive candidates generated by interest groups or organizations from opposite ends of the political spectrum.

II. HOW TOURNAMENTS CAN AID SUPREME COURT SELECTION

While a tournament is unlikely to reduce the role of politics in Supreme Court selection significantly or have a dramatic impact on the behavior of the parties in the confirmation process, it may nevertheless be useful to help correct some of the process’s pathologies. Presidents may find it helpful to conduct a credible search that at least appears to reduce the level of “politics.” The Senate may use tournaments to construct a meaningful “advice” function that provides incentives for the President to consult with Senate members prior to selecting a nominee.60 Interest groups can use the tournament to compete with the decidedly opaque American Bar Association (ABA) evaluation process. Finally, a tournament, if used by interest groups and the media, could provide a useful way of evaluating

---

60. See U.S. Const. art. II, § 2, cl. 2 (establishing the Senate’s role in Supreme Court appointments).
Supreme Court nominees without resorting to crude labels like “liberal” and “conservative” or “activist” and “strict constructionist.”

A. Improving the President’s Selection Process

Despite the considerable discretion Presidents have in selecting nominees, they are inevitably constrained by political realities. As David Yalof has demonstrated, this fact (as well as differences in personality) has caused Presidents’ approaches to Supreme Court selection to vary. Insensitivity to those realities can create problems not only for the nominee but also for the President, who can find that his “personal prestige and ability to influence other public matters suffer when the selection of a Supreme Court nominee fails to adequately account for the hostile forces at work in the immediate political context.” If recent struggles over lower court judges are a guide, the political landscape facing the current President and his nominee(s) will be a difficult one indeed. President Bush would do well to recognize and accept the political environment in which his Administration must operate; here, perhaps, the tournament might identify a nominee who would have a better chance at confirmation than one chosen through some other selection process. “Successful confirmation politics,” Yalof sensibly concluded, “often depends on whether the president has made astute selection decisions during [the] earlier stages of the appointment process.”

Yalof offered three models of presidential selection: (1) open selection, where decisions are made after vacancy; (2) single-candidate focused arrangements, where the President has someone in mind for the next vacancy; and (3) criteria-driven frameworks for selection, where “the president and his advisors set forth specific criteria to be met by prospective nominees.” Half of the modern nominees discussed in his book were selected by the third process. Yalof concluded that this approach had the best potential for permitting

61. See YALOF, supra note 22, at 4.
62. See id. at 4-5.
63. Id. at 4.
64. Id. at 168.
65. Id. at 6-7.
66. Id. at 177.
67. Id. at 177.
Presidents to further the interests of their administration, though
criteria-driven searches often require delegation to subordinates who
might become involved in turf wars with each other.68 While Yalof
was loath to offer “generalizations about relevant do’s and don’t’s in
Supreme Court recruitment,”69 he seemed to offer qualified approval
of the criteria-driven process, as long as one “eminently qualified
presidential advisor” with political and legal savvy was in charge and
the criteria chosen were not “overly restrictive,” which would pre-
serve flexibility and prevent “the quality of selection outcomes [from
being forced] down to its lowest common denominator.”70

Based on Yalof’s study and his tentative conclusions, Choi and
Gulati’s tournament could be of considerable help to an administra-
tion’s pursuit of a successful nomination and confirmation. Commit-
ting resources to identify potential nominees before a vacancy occurs
permits Presidents to consider different strategies in the absence of
an immediate need to make a nomination. Moreover, it permits a
President to signal to other players that the selection of Court per-
sonnel is important to the administration. Running a tournament
based on ex ante criteria could enable the administration to make
credible claims of a merit-based approach, even if it is unlikely that
the administration’s opponents will agree with the criteria that are
used.71 The appearance of merit-based selection (even when con-
ducted within ideological constraints) might suffice to shift the bur-
den to opponents of the nominee to justify their opposition in terms
of other criteria they would be obliged to disclose, as Choi and Gulati
suggest, thus giving a President (all things being equal) a strategic
advantage at the outset of the nomination.

The President may also be able to use a criteria-based tournament
to offer key Senators a role in the selection process. Asking Senators
to give names is always tricky: failure to choose a Senator’s preferred
nominee can create an enemy (even within one’s own party), and
given the present confirmation climate, Presidents need all the
friends they can get in the Senate. Alienating Senators is as impolitic
as it is unnecessary. But imagine a President inviting Senators to of-
ferr criteria by which future selections ought to be made. Co-opting
Senators by offering them the opportunity to construct the tourna-
ment both signals that the President respects the Senate’s “advice”
role and helps spread responsibility. Candidates who emerge from a
tournament that Senators had a role in constructing might have an

68. Id. at 172-80.
69. Id. at 186.
70. Id. at 187.
71. See supra Part I.D.
easier time than those emerging from a process that did not involve the Senate at all.

However, unless a future President tries to employ a tournament, we will have no way to test the possibilities presented above. As commentators have made clear, a good part of a President’s success in nominating a Supreme Court Justice has less to do with the nominee and more to do with the perceived power of the nominating President. Moreover, a President may find his efforts to introduce objective criteria stymied by bureaucratic infighting among his own advisors, by his own political commitments, or by the political context of the time. Perhaps Senators would not want to collaborate with the President, realizing that such collaboration may constrain their own actions in the future with regard to a particular nominee. Finally, any role for a tournament would depend at least in part on the level of a President’s interest in the Supreme Court and its personnel; history shows that this cannot be taken for granted.

B. Creating a Senate “Advice” Mechanism

Charles Black once concluded ruefully that any sort of formal Senate advice mechanism no longer exists. But there is no reason to think that it is too late in the day to create one; Choi and Gulati’s tournament may offer a way to do so. Senators may lobby the White House to have their favored criteria included in an administration-run tournament, which an administration may favor over the submission of specific names. Alternatively, Senators may wish to construct their own tournament and publicize the names of the winners to pressure the White House to take account of their wishes.

Perhaps Senators could even offer an administration a deal: allow us to participate in some way, and we will guarantee “fast-track”

72. See Abraham, supra note 22, at 28 (listing “opposition to the nominating president” as one of eight historical reasons for rejection of Supreme Court nominees); see also John Anthony Maltese, The Selling of Supreme Court Nominees 128-40 (1995) (discussing the “wide range of resources for screening potential nominees and generating support for them” available to modern Presidents); Segal, Cover & Cameron, supra note 44, at 506 (noting the likelihood of a nominee’s defeat “when . . . the political environment is hostile to the president”).

73. See, e.g., Yalof, supra note 22, at 134-35 (“Reagan’s hands-off management style ultimately encouraged a disproportionate amount of conflict among advisors within his administration, each clamoring for influence over the final selection outcome.”).

74. See, e.g., Michael J. Gerhardt, The Federal Appointments Process: A Constitutional and Historical Analysis (2000) (discussing lack of interest in judicial nominees on the part of the Clinton Administration); Yalof, supra note 22, at 40 (noting Truman’s “limited interest in most of the Court’s docket”).

75. Charles L. Black, Jr., A Note on Senatorial Consideration of Supreme Court Nominees, 79 YALE L.J. 657, 659 (1970) (“Procedurally, the stage of ‘advice’ has been short-circuited.”). I have argued that the “blue slip” that home-state Senators must submit to permit a judicial nomination to go forward can be seen as an effort to encourage Presidents to consult Senators prior to a nomination. Denning, Blue Slip, supra note 5, at 91-97.
consideration of nominees, including expedited hearings and a quick vote, as Glenn Reynolds once suggested. Under Reynolds’s proposal, the Senate would proffer a list of candidates to the President.

This list would constitute the “advice” portion of the Senate’s constitutional role. The President could then do one of two things—she could select a nominee from the list, who would be presumed competent based on the Senate’s earlier screening and would be given approval according to some sort of accelerated procedure . . . , or she could select someone not on the list, in which case the confirmation process would take place as usual.

Reynolds’s brief proposal did not include a discussion of the criteria the Senate should use to compile the initial list, but the procedure he outlined neatly accommodates Choi and Gulati’s tournament. If Senators could convince the media, interest groups, and the public that their selection process was fair and nonpartisan by citing their use of “neutral” criteria, then Choi and Gulati’s prediction that the burden would shift to the President to explain his choice if it was not on the Senate’s list could prove accurate.

It would be ideal if, in either event, Senators from both parties—perhaps those on the Senate Judiciary Committee—worked together to fashion criteria, or to run the tournament, or both. Wishful thinking? Perhaps, but such collective responsibility may give Senators of all parties reprieve from interest groups that pressure them to support or oppose particular nominees. Senators themselves have, of late, expressed frustration with the present process, and if the recent wrangling over courts of appeals judges are any indication, the next Supreme Court vacancy will produce an unprecedented bloodletting. Because Senators and the President realize this, perhaps now is an ideal time to propose an alternative to the mutually assured destruction of the current process.

77. Id. at 1580 (footnote omitted).
79. We should probably take these expressions of frustration with a grain of salt. As Reynolds reminded us, [I]t may be that the process as it now exists suits a lot of people more than they want to admit. Presidents and presidential candidates are able to promise to appoint ideologues of one stripe or another to the Court—and, if they can slip the ideologues past the Senate, to fulfill their promises. Senators are able to posture and curry favor with interest groups . . . . And those interest groups themselves are given a potent tool for currying publicity and raising
C. Ending the ABA’s Monopoly on Candidate Vetting

In the 1940s, the ABA began evaluating nominees to the federal bench. Presidents and Senators helped the ABA attain a role in the selection and confirmation of judges that was unique among interest groups. Acting on long-standing complaints by conservatives that the ABA was biased against conservative judges, the Bush Administration and Senate Republicans ended the ABA’s official role in vetting judicial nominees. The Administration’s action was criticized, and if Democrats retake the Senate (as they did briefly in 2001), the ABA may be restored to its prominent position.80 For those administrations unhappy with the ABA, because of suspicions about its agenda, dislike of the committee’s penchant for secrecy, or its elite makeup, Choi and Gulati’s tournament offers an alternative that does not leave an administration open to charges—like those leveled at the current Bush Administration—of letting ideology drive selection to the exclusion of other considerations.

The ABA committee in charge of judicial selection is made up of fifteen lawyers81 who rate nominees according to “professional qualifications—integrity, professional competence and judicial temperament.”82 As Choi and Gulati note, the ABA’s vetting process in no way approaches a tournament because they evaluate only candidates who are nominated, the standards used are not objective or easily quantifiable, and the process (including the makeup of the committee) is not transparent.83 Though the evidence of political bias is mixed, the ABA has certainly been perceived as biased against judges nominated by Republican Presidents,84 a perception no doubt...
reinforced by the ABA’s public stands on controversial legal and social issues.85

Critics have charged that there is no good reason to privilege the ABA over other interest groups. Choi and Gulati’s proposal now furnishes other groups—whatever their interest in the selection process—with a tool for publicizing their own evaluations of judicial nominees, or for suggesting nominees who “win” their tournament. At the very least, perhaps the introduction of some competition into the nominee evaluation game would put pressure on the ABA to open up its process and refine the notoriously vague criteria that its evaluation committee uses. If such alternative evaluations were done conscientiously and developed a reputation for fairness and transparency, administrations could choose to forego the ABA evaluations, yet they would be insulated from the charge that they were trying to shield their candidates from scrutiny or were pursuing a narrow ideological agenda. Even if having competing evaluations is confusing because of the proliferation of information, as I suggest above, the net result would be more information that members of the public could either sort through themselves or, as is now the case, rely on specialists and the media to digest and report.

D. Creating a New Vocabulary for Debating Nominations

The last point—that to the extent the public has an opinion at all on judicial nominees, it is one inevitably shaped by information filtered through interest groups and media outlets—is another problem for which Choi and Gulati’s tournament proposal offers a potential solution. When candidates are discussed, they are inevitably pigeonholed as liberal or conservative, as activists or practitioners of judicial restraint, or as pro-business or pro-civil rights. Where such labels are affixed to sitting judges (as Choi and Gulati point out, the next Supreme Court nominee will likely be a sitting judge),86 they tell almost nothing about the judge’s record, experience, or expertise and can, in fact, be extremely misleading.

A particularly disturbing trend among newspaper editorial boards is the tendency to condemn a judge because she “ruled against” a particular group (civil rights litigants, businesses, the state in criminal matters, and so on) a certain percent of the time.87 Such statistics

86. See Choi & Gulati, supra note 1, at 303 (“The norm today appears to be that a candidate for the Supreme Court must first sit on a federal circuit court of appeals before she may be considered for a seat on the Court.”).
tell little about either a judge’s abilities or her ideological disposition. They ignore entirely the underlying merits of the cases filed. Moreover, with lower court judges, a particular type of claim may be foreclosed by precedent. Problems are compounded when one tries to assign responsibility for the decision of multimember panels.

Using the number of times a judge is overruled or reversed is similarly unhelpful. A judge with a heavy caseload might be expected to have a higher number of reversals, which if converted to a percentage might be quite small. Appeals court judges are vulnerable to the whims of the U.S. Supreme Court. An appeals court could make a good-faith effort to apply Court precedent, only to have the Court adopt an equally plausible interpretation of its own precedent or repudiate its prior decisions and proceed in an entirely new direction.88

Senate confirmation hearings do little to improve upon the media characterizations. My colleague Bill Ross has demonstrated that the often jejune questions posed to nominees by members of the Senate Judiciary Committee—Senators presumably having a high level of interest and expertise in legal matters—show Senators wedded to the same reductionist dichotomies as the popular media. These questions in turn invite equally banal responses from the nominees themselves, who can perhaps be forgiven their reticence given the nature of the process.89

Choi and Gulati ask, “[W]hat if journalists today were to begin looking at the citation counts or publication rates of the Republican candidates for the court . . . ?”90 They predict that “[t]he tournament will have begun,” because comparisons would be made between nominees and other judges, criticisms would be made about the criteria used, and new metrics would be put forward.91 Even if the result would be competing tournaments, with The New York Times running one and The Wall Street Journal running another, they, like competing public opinion polls, would have to disclose their criteria and would be open to criticism.

Such efforts would, one hopes, shape the debate so that the next Supreme Court nominee would not be asked about her favorite Jus-
tice, or what Supreme Court decisions she disagreed with most,92 but perhaps, instead, about judicial independence, approaches to application of precedent, and prior decisions. The tournaments, moreover, might generate information that could be used to educate Senators, who are often satisfied with bland and uninformative assurances about fidelity to precedent, affinity for Justice John Marshall Harlan, abhorrence of *Lochner*93 (or *Korematsu*94 or *Dred Scott*95), and vigorous renunciations of judicial activism in any form. Similarly, solid information about a nominee that tournaments might provide could inoculate the nominee against charges of being “out of the mainstream” if the nominee is shown “objectively” to stack up well against her colleagues on the bench. Either outcome would be a welcome improvement over the present system.96

Finally, it is worth noting that Senators have not yet decided whether it is publicly acceptable to oppose an otherwise-qualified Justice solely on the basis of ideology (save for nominees deemed to be “extremists”). It is generally accepted, however, that opposition based on ethical lapses or questions about a nominee’s integrity is perfectly acceptable. The process thus creates an incentive for Senators who might otherwise oppose a nominee based on her ideology to play ethical gotcha games with nominees to justify a “no” vote. Similarly, since “extremists” are fair game as well, as I alluded to above, nominees sometimes find their records twisted to support an outcome made on other grounds. To the extent the tournament caught on, it might help expand the range of acceptable reasons to oppose a President’s nominees beyond impugning a nominee’s integrity or assassinating her character.97

III. CONCLUSION

It is probably impossible to remove politics from the Supreme Court selection process. Even if it were possible, I am not sure it would be desirable, given the tremendous power Supreme Court Justices potentially wield. That said, I do not accept that we are stuck

---

96. See also Gerhardt, supra note 18, at 366 (noting that the media’s move from “reporting ‘hard’ news or facts and figures to reporting ‘soft’ news or speculation and commentary,” due in part to the emergence of the twenty-four hour news cycle, has “put enormous pressure on newspapers and television reporters to emphasize scandal” with the result being that “[t]he media . . . prefers to stick to simple labels of ‘liberal’ and ‘conservative’.”).
97. Cf. ROBERT A. KATZMANN, COURTS AND CONGRESS 39 (1997) (“It may be that open and serious discussion of the nominee’s values, approach to the law and to decisionmaking, and declared policy preferences may have the salutary effect of reducing the temptation to do battle in highly personal terms.”).
with the pathogenic confirmation process to which we have become accustomed over the last twenty years. Politics and ideology have been employed as bases for opposition to Supreme Court nominees for nearly as long as the Court has existed. Yet only recently, it seems, has “war” become an appropriate metaphor for the process. Therefore, I applaud Choi and Gulati for rejecting the fatalism that has crept into discussions about the confirmation process and offering an alternative to it.

While not the depoliticized, objective, merit-machine academics might wish for, I believe that their tournament, or something like it, might be used to improve our present process. It offers the possibility for cooperation on Supreme Court appointments between the President and the Senate without either seeming to lose face or cede power. It offers a framework in which the Senate could revive a meaningful advice function. (If it served as even the beginning of a conversation between a presidential administration and Senators of both parties or simply spurred a bipartisan discussion among Senators about confirmation criteria, as it has among the participants in this Symposium, surely it was worth proposing.) Tournaments may create competition to produce the best measure for candidates, and may even spur long-time players, like the ABA, to refine their criteria in order to retain influence. Finally, it may help reframe the terms on which debates over nominees are conducted.

In the end, nothing may come of Choi and Gulati’s proposal. Perhaps, as Glenn Reynolds suggested, the present system suits the parties just fine, despite their public protestations. But if we are still debating the confirmation wars ten years from now, academics should remind Presidents, Senators, interest groups, and members of the media that it was not for lack of alternatives that the process did not improve.

98. See supra note 79 and accompanying text.
THE ROLE OF QUALIFICATIONS IN THE CONFIRMATION OF NOMINEES TO THE U.S. SUPREME COURT

LEE EPSTEIN,* JEFFREY A. SEGAL,** NANCY STAUDT*** AND RENÉ LINDSTÄDT****

In light of concerns that politics, philosophy, and ideology now dominate the federal judicial appointment process—a process that many claim should emphasize ethics, competence, and integrity—scholars have offered a range of proposals. A considerable number, though, aim to compel elected actors to focus on the candidates' qualifications rather than on their political preferences.

Without taking a normative position on these sorts of proposals, we demonstrate empirically that the process leading to the appointment of (at least) Supreme Court Justices may not be the “mess” that the proposals suggest. While it is true that U.S. Senators are more likely to cast votes for nominees who are ideologically proximate to them, qualifications also play a significant role in accounting for the choices Senators make.

I. INTRODUCTION .......................................................... 1146
II. THE CONFIRMATION OF SUPREME COURT JUSTICES: POLITICS VERSUS QUALIFICATIONS .......................................................... 1149
III. QUALIFICATIONS AND THE CONFIRMATION OF SUPREME COURT NOMINEES: OUR STUDY .......................................................... 1153
   A. The Dependent Variable: Confirmation Votes .......................................................... 1155
   B. An Independent Variable of Primary Interest: Qualifications .......................................................... 1157
   C. Ideology .................................................................................................................. 1161
      1. The Ideology of the Nominee .................................................................................. 1161
      2. The Ideology of Senators ...................................................................................... 1163
      3. Comparing the Ideology of Nominees and Senators .............................................. 1165
   D. Control Variables ...................................................................................................... 1166
   E. Summary of the Model .............................................................................................. 1167
IV. RESULTS OF THE ANALYSES .............................................................................. 1168
V. DISCUSSION .............................................................................................................. 1173

* Edward Mallinckrodt Distinguished University Professor of Political Science and Professor of Law at Washington University in St. Louis (http://epstein.wustl.edu/) (last visited Feb. 7, 2005). We are grateful to the National Science Foundation for supporting our research and Andrew D. Martin for offering useful comments. We used R (http://www.R-project.org) (last visited Feb. 7, 2005) and Stata to conduct the analyses and to generate the graphs presented in this Essay. The project’s website (http://epstein.wustl.edu/research/qualified.html) (last visited Feb. 7, 2005) houses the database. Please send all correspondence to Lee Epstein. Email: epstein@wustl.edu. Post: Department of Political Science, Washington University, CB 1063, 1 Brookings Drive, St. Louis, MO 63130.
** Professor of Political Science at Stony Brook University (http://www.sunysb.edu/polsci/segal/indexbody.html) (last visited Feb. 7, 2005).
*** Professor of Law at Washington University in St. Louis (http://law.wustl.edu/Academics/Faculty/Staudt/) (last visited Feb. 7, 2005).
I. INTRODUCTION

In recent testimony before the U.S. Senate, the legal scholar Ronald D. Rotunda declared that:

Our judicial system is at the top of the food chain, and that is a good reason to leave well enough alone. Given the fact that the Senate has been confirming federal judges for years, and the product is admired around the world, one wonders why we should think of changing the way the Senate confirms.¹

Without doubt, the vast majority of Rotunda’s colleagues—not to mention many members of the Senate—would disagree. For at least two decades now, they have deemed the federal confirmation process a “mess,”² “abysmal,”³ “broken,”⁴ “going in the wrong direction,”⁵ and downright “disorderly, contentious, and unpredictable.”⁶ Of course, pinpointing the precise cause of the “problem” has generated its own share of controversies.⁶ But there does seem to be general agreement that politics, philosophy, and ideology now dominate a process that should emphasize ethics, integrity, and competence—and as a result, the quality of our nation’s judiciary, along with its independence, has suffered if not markedly declined.⁷

---

⁵ Mark Silverstein, Judicious Choices: The New Politics of Supreme Court Nominations 6 (1994).
⁶ E.g., Herman Schwartz, Packing the Courts 166 (1988) (claiming that the “crusade” of conservative interests to “pack the courts with ideological zealots has induced more rather than less partisanship” in the selection process); Silverstein, supra note 5, at 6 (asserting that the current state of the confirmation process reflects “profound changes in American politics and institutions,” not the least of which is the “heightened activism of the modern federal judiciary”); Ruth Bader Ginsburg, Confirming Supreme Court Justices: Thoughts on the Second Opinion Rendered by the Senate, 1988 U. ILL. L. REV. 101, 117 (arguing that the problem is with the President’s failure to “seek[] more ‘advice’ from the Senate prior to a nomination,” which would keep the media and interest group campaigns “within more tolerable bounds”); Jeffrey K. Tulis, Constitutional Abdication: The Senate, the President, and Appointments to the Supreme Court, 47 CASE W. RES. L. REV. 1331, 1331 (1997) (suggesting that the problem lies with “[t]he altered relation of President and Congress in the appointment of justices of the Supreme Court”).
⁷ See, e.g., Carter, supra note 2; Choi & Gulati, supra note 3. On the other hand, we hasten to note, there are at least a handful of scholars who disagree with the view that politics is necessarily a danger to be avoided at all costs. See, e.g., Erwin Chemerinsky, Ideology and the Selection of Federal Judges, 36 U.C. DAVIS L. REV. 619, 620 (2003) (assert-
In light of these concerns, policymakers, social scientists, and legal academics have offered a range of proposals aimed at refocusing the spotlight on a candidate’s qualifications for office, rather than his or her ideology. Some have suggested, for example, that while the President should be free to consider any criteria he deems relevant, the Senate ought to eschew detailed examinations of the nominees’ politics and philosophy. As Professor Douglas W. Kmiec recently told the Senate:

My proposition is simple: the proper Senate inquiry of a judicial candidate is demeanor, integrity, legal competence, and fidelity to the rule of law. It is not partisanship or policy agreement. While textually the Senate is free to inquire and to reject a nominee on any ground—even a highly political, constitutionally problematic one like the nominee’s views on outcomes in specific cases—it should not do so. Undertaking to make nominees carry a type of political burden of proof will over time merely invite a subservience of mind and personality that is contrary to an independent judiciary.

Others cast only a partial concurrence, asserting that both the Senate and President should focus exclusively or almost exclusively on a nominee’s “objective” or “technical” qualifications. Some even
go so far as to propose a “tournament” among judges of the U.S. courts of appeals to ensure the selection of “effective” Justices,\textsuperscript{12} while simultaneously reducing the “level of partisan bickering.”\textsuperscript{13}

Without taking a normative position on these sorts of proposals—or registering our complete agreement with Rotunda’s claim that “[o]ur judicial system is at the top of the food chain, and that is a good reason to leave well enough alone”\textsuperscript{14}—we demonstrate empirically that the process leading to the appointment of (at least) Supreme Court Justices may not be the “mess” they suggest. In general, we find that while it is in fact the case that U.S. Senators are more likely to cast votes for nominees who are ideologically proximate to them, it is also true that the nominees’ qualifications play a significant role in accounting for the choices Senators make.\textsuperscript{15} More specifically, we show that qualifications have a significant impact on Senators who are ideologically distant from a nominee. That is, while Senators may very well support a politically akin candidate regardless of his or her professional merit—they also will cast a yea vote for a high-quality nominee regardless of his or her ideology.

We develop these findings in four steps. In Part II, we briefly consider allegations about the growing role of politics and the concom-
tant declining role of qualifications in the appointment of Justices. Part III describes the statistical model we deploy to assess the role of ideology versus qualifications in the Senate’s decision over Supreme Court nominees. Because the measures we invoke to animate the concepts of “ideology” and “qualifications” are crucial to the credibility of our modeling exercise, we describe them in some detail. Next, in Part IV, we turn to the results yielded by our statistical analysis; in particular, we examine the substantive effects of ideology and of qualifications on the votes of individual Senators. Since that analysis, as we foreshadow above, underscores the importance of qualifications for a seat on the Supreme Court, our findings deserve some attention in light of existing proposals to change how the United States appoints its judges. We take up this matter in Part V.

II. THE CONFIRMATION OF SUPREME COURT JUSTICES: POLITICS VERSUS QUALIFICATIONS

Academics and policymakers alike have expressed a number of concerns over the process by which Supreme Court Justices attain their seats on the bench. But one concern rising above nearly all others centers on the increasingly politicized nature of the process. In some of this commentary, the focus is on the role of organized interests and the media. Silverstein summarizes their influence by stating that “[t]he harsh reality . . . that modern interest group and media politics shape the selection of judges to our highest courts . . . has provoked a good deal of concern on the part of politician and citizen alike, and calls for the reform of the process of ‘advice and consent’ are frequently heard.” In other investigations elected actors move to center stage. Schwartz, for example, has argued that “[a]lthough partisan politics are inevitable to some extent, such considerations . . . [until the 1980s] played a very minor role in lower court appointments and only a slightly greater role in Supreme Court nominations.” Carter, too, agrees that the process is now overtly and overly politicized, but he emphasizes ideology rather than “party labels”: “Litmus tests,” he writes, “became far more important [during the Reagan and Bush I administrations]—and far more consistent—than at any time in the past.”

Has the process, in fact, grown more political in any or all of these ways? If we focus exclusively on the amount of media and interest group attention to nominations, the answer is undoubtedly yes. Consider, for example, Figure 1, infra, which shows the number of groups supplying

---

16. See, e.g., ABRAHAM, supra note 11; CARTER, supra note 2; CHOI & GULATI, supra note 3; Ginsburg, supra note 6.
17. SILVERSTEIN, supra note 5, at 164.
18. See SCHWARTZ, supra note 6.
19. CARTER, supra note 2, at 71.
oral or written testimony for or against each nominee since Earl Warren.\textsuperscript{20} Even though the most recent nominees (Justices Ginsburg and Breyer) failed to generate substantial interest group opposition (or support) relative to other recent nominees, the data overall evince a clear upward trend: A simple bivariate regression shows that with each passing nomination since Warren’s, 1.23 more groups participated.\textsuperscript{21}

\begin{figure}
\centering
\caption{Number of interest groups presenting oral or written testimony for or against each nominee from Earl Warren to Ruth Bader Ginsburg.\textsuperscript{22}}
\end{figure}

\begin{table}[h]
\centering
\begin{tabular}{|l|l|l|}
\hline
\textbf{VARIABLE} & \textbf{COEFFICIENT} & \textbf{(STANDARD ERROR)} \\
\hline
Counter & 1.232\textsuperscript{**} & (0.282) \\
Intercept & -4.557 & (4.351) \\
\hline
\end{tabular}
\end{table}

\textsuperscript{20} Lee Epstein et al., The Supreme Court Compendium 361 (3d ed. 2003).
\textsuperscript{21} Although we should not put much weight on a bivariate regression, the results are as follows (** indicates $p \leq .001$):
\textsuperscript{22} Id. Warren was nominated in 1953 and confirmed in 1954; Kennedy was nominated in 1987 and confirmed in 1988. Note that we do not include Homer Thornberry or Douglas Ginsburg because those nominations were withdrawn before any Senate action. The Fortas nomination (for Chief Justice in 1968) was also withdrawn, but only after a Senate vote to end a filibuster failed to receive the necessary two-thirds majority.
However intriguing the data may be, it represents but the tip of the iceberg. The data captures only activity that occurs in public view—and on the floor of the Senate—when we know interest groups regularly lobby behind the scenes, whether in the corridors of Congress or on the pages of The New York Times. But even with this limitation it would be hard to argue that the data displayed in Figure 1 fails to support Caldeira and Wright’s astute observation:

[T]he selection of [Supreme Court Justices]—once a “cozy triangle” of senators, the executive branch, and the bar—has become a major arena for the participation of interest groups . . . . What is more, despite changes in administration, the broad participation of organized interests and the battle lines drawn in the 1980s over the politics of judicial nominations persist.

Conveying a similar message about the increasingly politicized environment surrounding the appointment process are the data in Figure 2, infra. There we illustrate the increase in media coverage of individual nominations over time, with the dots indicating the precise magnitude of the growth (for example, Justice Ginsburg’s appointment in 1994 generated 49 more stories [in The New York Times and Time magazine] than did her predecessor, Byron White’s about three decades earlier, in 1962). Note that in every instance the successor nominees received more coverage than their predecessors and that overall, The New York Times and Time magazine published 358 more stories about the post-1980 nominees, for a mean change of 51.14.

---

23. Indeed, one study estimates that at least 150 organizations participated in one way or another in the Bork nomination, David Austen-Smith & John R. Wright, Counteractive Lobbying, 38 AM. J. POL. SCI. 25, 35 (1994), while just thirty-eight provided testimony.

24. Gregory A. Caldeira & John R. Wright, Lobbying for Justice: The Rise of Organized Conflict in the Politics of Federal Judgships, in CONTEMPLATING COURTS 44-45 (Lee Epstein ed., 1995). Nonetheless, it is worth pointing out that while interest group participation in nominations is on the upswing, it is “not an entirely new phenomenon,” as Yalof points out. “Organized interests,” he continues, “figured significantly in defeating Stanley Matthews’s nomination to the Court in 1881. Almost fifty years later, an unlikely coalition of labor interests and civil rights groups joined together to defeat the nomination of John Parker.” DAVID ALISTAIR YALOF, PURSUIT OF JUSTICES 16 (1999) (footnote omitted); see also Caldeira & Wright, supra, at 47-50.
These data lend some support to those who argue that the environment surrounding Senate contemplation of Supreme Court nominees has grown increasingly political and highly charged. But the more relevant question is the extent to which U.S. Senators are political in response. Do Senators’ votes, in other words, attend to ideological and partisan-political concerns rather than to a candidate’s qualifications to serve on the high Court?

The answer, according to much of the existing commentary, is yes. As Watson and Stookey write, a common complaint about the process is that “voting of the Senate has reflected political motives and ignored whether in fact the nominee is qualified.” Carter, the author of *The Confirmation Mess*, surely concurs, as does Silverstein:

> It is difficult, if not impossible, to imagine prominent members of one political party today championing the nomination of a member

26. Exceptions here include Cameron et al., *supra* note 15; Segal et al., *supra* note 15. In Part IV we refine and update the Cameron analysis.
of the opposition to the Supreme Court, and the notion that a
president might be constrained to seek only nominees of stature
and prominence is simply too fanciful to be seriously entertained.
That a Democratic nominee with the long public career and the
controversial publications of a [Benjamin] Cardozo could be nomi-
nated by a Republican president and confirmed by a voice vote of a
Republican-controlled Senate, all within a period of ten days, is it-
self beyond modern comprehension.29

This is but a sampling of the conventional wisdom regarding the
triumph of politics over credentials in the confirmation process;
other exemplars would hardly be difficult to unearth.30 Nor, we
hasten to note, would it be difficult to locate social science evi-
dence supporting these beliefs. From their analyses of seven
nominations—Fortas (1968), Haynsworth, Carswell, Rehnquist
(1971), Rehnquist (1986), Bork, and Thomas—Watson and Stookey
claim that they can accurately predict 81.43% (n = 643) of the
approximately 700 votes cast based solely on the ideology and politi-
cal party affiliation of the Senators.31 Massaro, in a study of the
failed Fortas, Haynsworth, and Carswell nominations, goes even
further—parsing the effect of the Senators’ ideology and partisan-
ship on their roll call votes. At the end of the exercise, Massaro
concludes that “[i]n all three nominations, ideology is indicated to
be a more convincing factor than party affiliation in explaining
Senate voting on the Fortas cloture roll call and on the
Haynsworth and Carswell nominations.”32

III. QUALIFICATIONS AND THE CONFIRMATION OF SUPREME COURT
NOMINEES: OUR STUDY

While these and other analyses seem to clinch the case—the con-
firmation of Supreme Court Justices is now much more about politics
and ideology than it is about integrity and ethics—they are not with-
out their flaws. Most relevant is that the authors select only those
nominations that are controversial, thereby begging the question of
what factors, including (the lack of) qualifications, lead to contro-
versy. It also is the case that the authors base their conclusions on

29. Silverstein, supra note 5, at 2.
30. See, e.g., Carter, supra note 2.
31. See Watson & Stookey, supra note 8, at 185, 190.
32. See John Massaro, SupremeLY Political: The Role of Ideology and
Presidential Management in Unsuccessful Supreme Court Nominations 15
(1990); see also David W. Rohde & Harold J. Spaeth, Supreme Court Decision
Making 106 (1976). Rohde and Spaeth conducted a similar investigation of the nomi-
nations of Fortas (for Chief Justice in 1968), Haynsworth, Carswell, and Rehnquist (for
associate in 1971) and reached the same general conclusion: “[I]t is the degree of liberalism
of a senator and not his party affiliation which is related to his voting on nominations.”
Id.
statistical models that take into account only politics and ideology; they fail to attend to qualifications, along with other factors that may affect the confirmation decision.\textsuperscript{33} As such, they cannot possibly assess the extent to which ideology, qualifications, both, or neither affect the votes cast by Senators. Making that assessment requires a consideration of all the relevant factors simultaneously in one statistical model.\textsuperscript{34}

To our knowledge, only one team of researchers has undertaken this task—Cameron, Cover, and Segal in their 1990 article appearing in *The American Political Science Review*\textsuperscript{35}—and it is that team’s lead that we follow here. Momentarily, we unveil our plan for so doing, that is, for refining and extending this classic study. Here and now, though, we simply want to outline the underlying logic of our (and the Cameron et al.) analysis. For starters, it is important to understand what we are seeking to explain, namely, the votes cast by individual Senators over Supreme Court nominees since Warren in 1953.\textsuperscript{36} These votes, in other words and in the parlance of social science, constitute the dependent variable in our study.

Second, we attempt to explain these votes via three sets of factors, or independent variables: qualifications,\textsuperscript{37} ideology,\textsuperscript{38} and control variables.\textsuperscript{39} The first two—qualifications and ideology—constitute the primary variables of interest; that is, we gear our study toward understanding whether nominees’ qualifications, their ideological proximity to Senators, neither, or both account for Senators’ votes. Control variables are those that we also think may affect confirmation votes, and therefore we must consider them to avoid “omitted variable bias.”\textsuperscript{40} Here those variables are, as they were in the Cameron et al. analysis, whether the President is “strong” and whether the Senator is of the same party as the President.\textsuperscript{41}

\textsuperscript{33} For other problems with these sorts of studies, see Cameron et al., supra note 15, at 526.
\textsuperscript{34} For more on this point, see Lee Epstein & Gary King, *The Rules of Inference*, 69 U. CHI. L. REV. 1, 76-80 (2002). See also infra note 40 and accompanying text.
\textsuperscript{35} Cameron et al., supra note 15; see also Segal et al., supra note 15. However, here we focus on extending and refining the earlier 1990 piece.
\textsuperscript{36} Warren was nominated in 1953 and confirmed in 1954. More generally, see infra Part III.A for more information on Senate votes over nominations to the Court.
\textsuperscript{37} See infra Part III.B.
\textsuperscript{38} See infra Part III.C.
\textsuperscript{39} See infra Part III.D.
\textsuperscript{40} Epstein & King, supra note 34; GARY KING ET AL., DESIGNING SOCIAL INQUIRY: SCIENTIFIC INFERENCE IN QUALITATIVE RESEARCH (1994). Omitted variable bias occurs when a statistical comparison excludes variables that are (a) known to affect the outcome and (b) correlated with the explanatory covariate of interest.
\textsuperscript{41} Cameron et al., supra note 15.
Finally, we require a statistical method that enables us to estimate the effect of these independent variables on the dependent variables (the votes of individual Senators). For the reasons we explain in Part IV, a maximum-likelihood probit model is ideally suited for our purposes.

With our general plan outlined, let us now turn to the task of fleshing it out by providing details on the dependent variable, the independent variables, and our statistical model. We then, in Part IV, turn to an inspection of the results yielded by our analysis.

A. The Dependent Variable: Confirmation Votes

The dependent variable of our study consists of the 2461 confirmation votes cast by individual Senators on the Supreme Court nominees from Earl Warren, in 1953, through Stephen G. Breyer, in 1994.\textsuperscript{42} Of the 2461 total votes, 15.35% ($n = 378$) were cast against the nominee and 84.65% ($n = 2084$) were cast in the nominee's favor.

Since our website houses the vote data, along with all other variables included in this study,\textsuperscript{43} we merely summarize, in Table 1, infra, the aggregate votes for and against each nominee. Note that most votes have not been close. Thus the mode of nay votes is zero. On the other hand, a great deal of variation exists from nominee to nominee, such that the number of nay votes on average is 14.81, with a relatively high standard deviation of 19.94.

\textsuperscript{42} Inter-University Consortium for Political and Social Research, United States Congressional Roll Call Voting Records (Study No. 4), at http://webapp.icpsr.umich.edu/cocon/ICPSR-STUDY00004.xml (last modified June 17, 2004).

\textsuperscript{43} Lee Epstein et al., The Role of Qualifications in the Confirmation of Nominees to the U.S. Supreme Court, at http://epstein.wustl.edu/research/qualified.html (last visited Oct. 20, 2004) (providing data for this Essay).
Table 1

SENATE VOTES ON SUPREME COURT NOMINEES, 1953-1994.44

<table>
<thead>
<tr>
<th>Nominee</th>
<th>Votes in Favor</th>
<th>Votes Opposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earl Warren (CJ)</td>
<td>96</td>
<td>0</td>
</tr>
<tr>
<td>John Marshall Harlan</td>
<td>71</td>
<td>11</td>
</tr>
<tr>
<td>William J. Brennan, Jr.</td>
<td>95</td>
<td>0</td>
</tr>
<tr>
<td>Charles Evans Whittaker</td>
<td>96</td>
<td>0</td>
</tr>
<tr>
<td>Potter Stewart</td>
<td>71</td>
<td>17</td>
</tr>
<tr>
<td>Byron Raymond White</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Arthur Joseph Goldberg</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Abe Fortas</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Thurgood Marshall</td>
<td>69</td>
<td>11</td>
</tr>
<tr>
<td>Abe Fortas (CJ)</td>
<td>43</td>
<td>44</td>
</tr>
<tr>
<td>Warren Earl Burger (CJ)</td>
<td>74</td>
<td>3</td>
</tr>
<tr>
<td>Clement Haynsworth, Jr.</td>
<td>45</td>
<td>55</td>
</tr>
<tr>
<td>G. Harrold Carswell</td>
<td>45</td>
<td>51</td>
</tr>
<tr>
<td>Harry A. Blackmun</td>
<td>94</td>
<td>0</td>
</tr>
<tr>
<td>Lewis F. Powell, Jr.</td>
<td>89</td>
<td>1</td>
</tr>
<tr>
<td>William H. Rehnquist</td>
<td>68</td>
<td>26</td>
</tr>
<tr>
<td>John Paul Stevens</td>
<td>98</td>
<td>0</td>
</tr>
<tr>
<td>Sandra Day O’Connor</td>
<td>99</td>
<td>0</td>
</tr>
<tr>
<td>William H. Rehnquist (CJ)</td>
<td>65</td>
<td>33</td>
</tr>
<tr>
<td>Antonin Scalia</td>
<td>98</td>
<td>0</td>
</tr>
<tr>
<td>Robert H. Bork</td>
<td>42</td>
<td>58</td>
</tr>
<tr>
<td>Anthony M. Kennedy</td>
<td>97</td>
<td>0</td>
</tr>
<tr>
<td>David H. Souter</td>
<td>90</td>
<td>9</td>
</tr>
<tr>
<td>Clarence Thomas</td>
<td>52</td>
<td>48</td>
</tr>
<tr>
<td>Ruth Bader Ginsburg</td>
<td>96</td>
<td>3</td>
</tr>
<tr>
<td>Stephen G. Breyer</td>
<td>87</td>
<td>9</td>
</tr>
</tbody>
</table>

44. The letters “CJ” indicate the President nominated the candidate for Chief Justice. The vote on Fortas for the Chief Justice position was on cloture and failed to receive the necessary two-thirds majority. Voice votes were taken on the nominations of Warren, Brennan, Whittaker, White, Goldberg, and Fortas (for an Associate Justice position). Id. The voice votes are treated as unanimous because there is no data on the counts. It is worth noting that we reran the model presented in Table 2, infra, at p. 1168, without the six “voice vote” nominees and all the variables continued to generate statistically significant (p < .05) coefficients (the p-value for Lack of Qualifications drops from .005 to .032, and the p-value for Ideological Distance drops from .010 to .012). Also worth noting is that Felice and Weisberg report that across four of the voice votes (Warren, Brennan, Fortas, and Goldberg) there was, in fact, opposition from a total of (at least) seven Senators. Felice & Weisberg, supra note 9, at 515. Six of these seven Senators’ votes were modified from yeas to nays (the seventh did not participate in the vote over the nominee he apparently opposed), and we reran the model depicted in Table 2. Once again, all the variables continue to generate statistically significant (p < .05) coefficients (the p-value for Lack of Qualifications moves from .005 to .002, and the p-value for Ideological Distance moves from .01 to ≤ .001).
B. An Independent Variable of Primary Interest: Qualifications

What the vote data reveals, to put it simply, is that while Senators overwhelmingly vote in favor of confirmation (84.65% of all votes are yea), we could hardly deem decisions over Supreme Court nominees consensual. In eighteen of the twenty-six cases, the nominee caused some degree of division among the Senators. This is important for our purposes, since without variation in votes, even limited variation, we would have little to explain; we might well conclude that the Senate merely supported (or opposed) the President’s choice.45

Since this is clearly not the case, let us now turn to the factors that may explain the variation we observe, beginning with qualifications. As we have noted throughout, many would agree with Choi and Gulati when they write:

[Discussion[s] over Supreme Court nominees are now] almost entirely political (focusing on litmus tests such as a candidate’s likely position on abortion). Occasionally, a nominee’s intellectual ability is mentioned, but this topic has time and time again been placed to the side in favor of a discussion of the nominee’s political beliefs.46

Choi and Gulati may be right in their characterization of the current state of appointment discourse, but does that necessarily mean that qualifications play an insignificant role (relative to other factors) in a Senator’s decision to vote for or against a candidate, especially when the Senator is ideologically distant from the candidate?

To address this question, we must develop a measure of “qualifications.” Given the disagreement among scholars and policymakers over the characteristics that make for a “qualified” nominee, this is no simple mission. Indeed, there are some who seem to argue that we ought to jettison it altogether. To them, attempts to devise “objective” indicators of qualifications or merit are “doomed to failure” because “[q]ualifications’ always have been and always will be defined politically.”47

We cannot say we disagree, but devising a measure of merit based on Senators’ (or even scholars’) colored definitions of merit is not our project. Rather, our goal is to tap into the Senators’ or—more precisely, assuming that Senators are oriented toward reelection—their constituents’ perceptions of whether a candidate is qualified or not. This requires us to locate a measure of qualifications from sources external to and independent of the Senate (and, of course, that is available and observable prior to its vote).

45. But we could not, nor should we, eliminate the possibility that the President nominated individuals he believed the Senate would confirm. See Byron J. Moraski & Charles R. Shipan, The Politics of Supreme Court Nominations: A Theory of Institutional Constraints and Choices, 43 AM. J. POL. SCI. 1069 (1999).
46. Choi & Gulati, supra note 3, at 300-01 (footnote omitted).
47. Watson & Stookey, supra note 8, at 219.
One measure that comes readily to mind is the Nominee ratings produced by the American Bar Association’s (ABA) Standing Committee on the Federal Judiciary. The ABA’s ratings are presumably extrinsic to individual Senators, are announced prior to the confirmation vote, and, according to the ABA, are “impartial evaluations of the integrity, professional competence and judicial temperament” that “[do] not consider a nominee’s philosophy or ideology.” However, they are also problematic in any number of ways. One is that the Committee’s rating system has fluctuated with time, and even within particular periods it has lacked consistency. For example, until 1970 it typically rated a candidate as simply “qualified” or “unqualified.” In 1963, however, it deemed Arthur Goldberg “highly acceptable,” but the ABA thought it inappropriate to proffer “an opinion to the degree of qualification.” Also problematic for our purposes are allegations that ABA ratings evince a (liberal) ideological bias; that bias may explain George W. Bush’s decision to end the Committee’s “semi-official” role in conducting pre-nomination evaluations of judicial candidates, a role it has played since the Eisenhower administration.

Given these concerns, we think it is best to eschew the ABA approach in favor of the one that Cameron and his colleagues developed and used in their study of Senate votes: a measure of qualifications derived from a content analysis of newspaper editorials written from the time of nomination by the President until the vote by the Senate. Specifically, Cameron and his colleagues selected four of the nation’s leading newspapers, two with a liberal outlook (The New York Times and The Washington Post) and two on the more conservative end (the Chicago Tribune and the Los Angeles Times) and identi-

49. ABRAHAM, supra note 11, at 25. For the language the ABA has used in its ratings, see EPSTEIN ET AL., supra note 20, at 359-60.
51. See Neil A. Lewis, Bar Association’s Role in Screening Federal Judges Is Reviewed, N.Y. Times, Mar. 20, 2001, at A16. The Bush administration has denied this charge, claiming instead, in a letter to the President of the ABA: The issue at hand . . . [is] whether the A.B.A. alone—out of the literally dozens of groups and many individuals who have a strong interest in the composition of the federal courts—should receive advance notice of the identities of potential nominees in order to render prenomination opinions on their fitness for judicial service.
Neil A. Lewis, White House Ends Bar Association’s Role in Screening Federal Judges, N.Y. Times, Mar. 23, 2001, at A13. Ultimately, the administration decided the ABA should not receive advance notice. Id.
52. See Cameron et al., supra note 15, at 529-30. The same research team also used the qualifications measure in Segal et al., supra note 15, and the scores are reported in EPSTEIN ET AL., supra note 20, at 361.
fied every editorial that offered an opinion on the candidate’s qualifications (from the nominations of Earl Warren through Anthony Kennedy). With the editorials in hand, Cameron and his colleagues coded their content on the basis of claims about the nominee’s acceptability from a professional standpoint; the research team then created a scale of qualifications for each nominee that ranges from one (most qualified) to zero (least qualified).

Following procedures set forth by Cameron and his colleagues, Segal updated their measure to include the four nominations subsequent to Anthony Kennedy, and we now display the results, the qualification scores for all nominees from Earl Warren to date, in Figure 3, infra. From the data it is easy to see why we and others find this measure so compelling: it seems to have a high degree of facial validity; that is, it appears to comport with our existing knowledge of the nominees. Note, for example, that it is Carswell, reckoned “mediocre” even by supporters, who receives the lowest score, while it is Kennedy, Ginsburg, Scalia, and several others—that is, candidates even would-be opponents admitted were qualified to serve— who received the highest score.

53. To provide an example (one of which comes from our updating of the Cameron et al. qualifications score), consider the following statement, which appeared in an editorial about Ruth Bader Ginsburg in The New York Times:

The bridge she builds to justices like John Harlan, who served from 1957 to 1971, is a reminder of the mediocrity of so many appointees of the Bush-Reagan years. Nominees chosen for ideology, or with sparse credentials out of political necessity, by increments have depressed the Court's performance, professional standing and fidelity to law. President Clinton's nominee brings a touch of class to the Supreme Court.

A Touch of Class for the Court, N.Y. Times, July 25, 1993, at E16. We coded this as a positive statement about Ginsburg's professional qualifications.

The following, also appearing in The New York Times, would be a negative statement about Clarence Thomas's professional qualifications:

Believe him or not, nothing in this bizarre episode enhances Judge Thomas's qualifications, which were slim to start. Believe him or not, his behavior on the witness stand does nothing to enhance those qualifications. Believe him or not, to confirm him is to gamble.

If Judge Thomas were a brilliant jurist, a Holmes or a Brandeis, the gamble might be justified. But Clarence Thomas offers no such brilliance, no basis for gambling with the public's confidence in, and the future of, American law.


54. In our analysis, we use a nominee's lack of qualifications, rather than his qualifications, as an independent variable. We derive the Lack-of-Qualification variable simply by subtracting the Qualification measure from one.

55. Recall Senator Roman Hruska's (infamous) defense of Judge Carswell: “Even if he were mediocre, there are a lot of mediocre judges and people and lawyers. They are entitled to a little representation, aren’t they, and a little chance?” LAWRENCE BAUM, THE SUPREME COURT 47 (1st ed. 1981).

56. Note that these three run the ideological gamut from conservative to relatively moderate to liberal, and yet all received the ABA’s highest endorsement: well qualified by a unanimous vote. See EPSTEIN ET AL., supra note 20, at 360.
Facial validity, though, is not the only reason why we invoke these measures in our quest to explain Senate votes for Supreme Court nominees. At least three others come to mind. First, the scores meet our original criteria of being external to the Senate (it is newspaper editors and not Senators from whom we derived the scores) and of being available and observable prior to the Senate’s vote. Second, the scores pass standard criteria for intercoder reliability: using $\pi$ as their index, Cameron and his colleagues report results of .87 ($p < .001$).58 Finally—and perhaps not so stunningly, given the range of newspapers consulted—the measure does not appear biased by ideology or political party; in other words, neither liberals nor democrats receive higher (or lower) ratings based solely on their policy preferences or partisanship.59

Despite these advantages, at least one scholar has critiqued the ap-
proach on the ground that it does not fit with his “conviction that the measure of characteristics such as ‘qualifications’ or even ‘ideology’ is never static but fluctuates over time in response to the political realities of the day.” But because the measure is indeed dynamic in this way—after all, its developers derived it from editorials contemporaneous with the nomination—this is actually yet another benefit of, rather than a fatal flaw in, our approach to assessing a candidate’s qualifications.

C. Ideology

“Qualifications” is one of our variables of chief concern; the other is ideology (or policy preferences). Specifically, in line with existing commentary, we expect that Senators are more likely to vote for nominees who are ideologically proximate to them (or their constituents)61 than they are for nominees who are ideologically distant, especially if the nominee is not particularly well qualified.

To assess this hypothesis we require not one but two measures of ideology—the candidate’s and the Senator’s—as well as a method for comparing the two so that we can calculate the distance between them. In what follows, we elaborate on these requirements and how we fulfilled them.

1. The Ideology of the Nominee

To assess the ideology of nominees, we must develop a measure that is independent of judicial behavior (or at least independent of behavior on the Supreme Court), that we can calculate for all nominees (thereby eliminating, for example, votes cast or opinions written as a lower court judge), and that we can observe prior to the Senate’s vote.

Since these are some of the very same criteria that guided our selection of a measure of qualifications, it will come as no surprise that our measure of ideology is quite similar to an ideological score developed by political scientists Segal and Cover from newspaper editorials written between the time of nomination to the Supreme Court and the Senate’s vote.62

The procedures used by Segal and Cover to analyze the editorials are virtually the same as those Cameron, Cover, and Segal invoked to create their measure of qualifications, but here, of course, they focused their

60. Silverstein, supra note 5, at 5.
61. We do not attempt to separate a Senator’s personal ideology from that of his or her constituents; rather, we rely on analyses of roll call votes in the form of NOMINATE scores. For an effort to separate Senate and constituent preferences, see Segal et al., supra note 15.
content analysis on ideology. As Segal and Cover tell it:

[W]e trained three students to code each paragraph [in the editorial] for political ideology. Paragraphs were coded as liberal, moderate, conservative, or not applicable. Liberal statements include (but are not limited to) those ascribing support for the rights of defendants in criminal cases, women and racial minorities in equality cases, and the individual against the government in privacy and First Amendment cases. Conservative statements are those with an opposite direction. Moderate statements include those that explicitly ascribe moderation to the nominees or those that ascribe both liberal and conservative values.63

Segal and Cover then measured judicial ideology by subtracting the fraction of paragraphs coded conservative from the fraction of paragraphs coded liberal and dividing by the total number of paragraphs coded liberal, conservative, and moderate. The resulting scale of ideology (or policy preferences) ranges from zero (unanimously conservative) to .50 (moderate) to one (unanimously liberal). Figure 4 displays the score for each post-1953 nominee.

**Figure 4**

**PERCEIVED IDEOLOGY OF SUPREME COURT NOMINEES, 1953-1994.** The scores range from zero (most conservative) to one (most liberal).

---

63. Segal & Cover, supra note 62, at 559.
64. The letters “CJ” indicate the President nominated the candidate for Chief Justice. 
See Epstein et al., supra note 20, at 361.
These ideological scores have developed quite a following in the social sciences, and it is not hard to see why. Just as the qualification score appears facially valid, so too do the ideological scores. To be sure, there are some exceptions (for example, Clarence Thomas seems more conservative than his score), but overall the measure comports with our impressions of those nominees who ascended to the Court. William Brennan and Thurgood Marshall, generally regarded as liberals, received scores of 1.00; Scalia and Rehnquist, generally regarded as conservatives, received scores of .00 and .05 respectively.65

For our purposes the scores’ degree of validity (and reliability)66 is important, but they are also of Court behavior, available for all our nominees, and observable prior to the Senate’s vote. In other words, the scores fulfill just about all the needs of our research.

2. The Ideology of Senators

Assessing the ideology of Senators is not a particularly challenging task. For over a decade now, social scientists have invoked NOMINATE scores or a variation of them (such as the ones we use here, Common Space Scores)67 to measure the ideology of Senators.68 These scores result from subjecting congressional roll call votes to a scaling algorithm designed to identify each Congressional member’s position in an ideological space.69 The first dimension coordinate (which we use here) typically picks up the liberal/conservative dimension of conflict in American politics and

---

65. In fact, scholars have found that the ideological scores provide a satisfactory predictor of judicial votes. See, e.g., Segal & Cover, supra note 62; Segal et al., supra note 62. Certainly they explain the votes in some issues better than they do in others. See Lee Epstein & Carol Mershon, Measuring Political Preferences, 40 AM. J. POL. SCI. 261 (1996). But overall, across a range of cases, they have above-threshold predictive power. For example, for the nominees in our study who attained seats on the Court, the correlation between the ideological scores and votes in civil liberties cases is .771 and for economic cases it is .620.

66. Using π, Cameron and his colleagues report reliability results of .72 (p < .001). Cameron et al., supra note 15, at 533.

67. Some scholars use the Americans for Democratic Action (ADA) vote scorecards to measure the ideology of Senators. But, because the ADA relies only on a subset (and a nonrandom one at that) of votes to compute its scores, we eschew this approach in favor of NOMINATE scores, and specifically the Common Space Scores. We have several reasons for taking this particular tack, most importantly because it provides us with scores for Presidents, and we ultimately deploy those scores to derive Common Space Scores for nominees. See infra Part III.C.3.


69. All votes with less than 97.5% agreement are scaled, and all members who voted at least twenty-five times in a given Congress are scaled.
ranges from negative one (most liberal) to one (most conservative).\footnote{We should note that by invoking these scores, we depart from Cameron and his colleagues who used the ADA vote scorecards. See supra note 67. In computing the ideological distance between the Senators and nominees, Cameron and his colleagues compared, on the same metric, the Segal and Cover scores (their and our measure of nominee ideology) and the ADA scores (their measure of Senate ideology). See Cameron et al., supra note 15, at 533. We, of course, use a different measure of Senate ideology (the Common Space Scores), and we take a different tack in computing the distance between Senators and nominees. See infra Part III.C.3. Nonetheless, we did replicate their model (using their measures but including the four most recent nominees) to assess the compatibility of our approach and theirs. No major distinctions arose, as even a quick glance at the table below and the one presented in Table 2, infra, at p. 1168, would reveal.}

Of course, our interest is in the ideology of the individual Senators who voted on Supreme Court nominees. But, for purposes of illustration, we provide, in Figure 5, infra, a glimpse of the ideology of the Senate during the nominations under study here. Specifically, we depict the NOMINATE Common Space Score of the median member of the Senate at the time of the confirmation proceedings from Earl Warren through Stephen Breyer.

\begin{tabular}{|l|c|c|}
\hline
\textbf{VARIABLE} & \textbf{COEFFICIENT} & \textbf{(STANDARD ERROR)} \\
\hline
Lack of Qualifications & -1.220** & (0.257) \\
Ideological Distance & -1.156** & (0.251) \\
Lack of Qualifications $\times$ Ideological Distance & -9.678** & (0.954) \\
Strong President & 1.020** & (0.117) \\
Same Party & 0.838** & (0.105) \\
Constant & 1.774** & (0.110) \\
\hline
\end{tabular}

\begin{itemize}
    \item N = 2451
    \item Log-likelihood $\chi^2(5) = -461.347$
    \item $\chi^2(5) = 1184.991$
\end{itemize}

Probit estimates of Senators’ votes on Supreme Court nominees, using the Cameron et al. approach (** indicates $p \leq .01$).
Interesting, of course, is the disparate ideological conditions prevailing at the various nominations. Note, for example, the relatively “friendly” environment surrounding the confirmation of Thurgood Marshall: a liberal nominee (Marshall attains a perfect one (liberal) score on Segal and Cover’s ideological scale) facing one of the most liberal Senates in our database (a median Common Space Score of -.1575). In direct contrast comes the conservative Robert Bork (a .10 on the Segal and Cover scale), who was forced to confront a relatively liberal Senate (a median Common Space Score of -.0665).\footnote{The median Senate first dimension Common Space Score over all the nominations is -.057, and the mean is -.033 with a standard deviation of .314. \textit{Id.}}

3. Comparing the Ideology of Nominees and Senators

Whether the hospitable political environment explains Marshall’s success and the relatively hostile environment explains Bork’s failure

\footnote{See Keith Poole, \textit{Common Space Data}, at http://www.voteview.com/readmeb.htm (last visited Feb. 7, 2005).}
remains an open question—and one that the forthcoming analyses are precisely designed to explain. But, before we turn to those analyses, we have several more steps to take, including among the most delicate: devising a method to compare the ideologies of Senators and nominees so that we can compute the distance between them. Only by deriving such a “distance” variable can we determine the extent to which politics enters into the decisions of individual Senators.

This is indeed a delicate task in light of the measures we employ to tap the preferences of these actors: the Segal and Cover scores for nominees and the Common Space Scores for Senators. Because these preference proxies are not directly comparable, merely subtracting one from the other will produce a problematic measure of the distance between them.

We solve this problem by making use of the underlying logic of the Cameron et al. analysis—which conceptualizes distance by squaring the difference between the ideology of the nominee and the ideology of the Senator—but we adapt it to our measurement strategy. Specifically, we arrive at a Common Space Score for each nominee so that we can directly calculate the distance between his or her ideology and that of the voting Senators. (We derived the score by applying a linear transformation, which we generated by regressing presidential Common Space Scores on the Segal-Cover scores.)

D. Control Variables

As we have stated throughout, “Ideology” and “Qualifications” constitute our chief independent variables, but there are two others that Cameron and his colleagues (and, indeed, any analysis of Senate voting over Court nominees ought) take into account: the “Strength of the President” and “Same Party.” The strength of the President cap-

---

73. Recall that Cameron and colleagues compared (on the same metric) the Segal and Cover scores (their and our measure of Court ideology) and the ADA scores (their measure of Senate ideology). Cameron et al., supra note 15, at 533; see also supra notes 67, 70. We would be in error to do the same given our assessment of Senate ideology via Common Space Scores.

74. The specific steps we took are as follows:
We began by estimating a simple OLS regression model with presidential ideology NOMINATE Common Space Scores as the dependent variable and the Segal and Cover scores as the only independent variable. We included only those Presidents whose party controlled the Senate at the time of confirmation, under the assumption that such Presidents are (relatively) unconstrained and thus able to appoint nominees who mirror their own ideology.

The OLS regression provided the following linear transformation to calculate Common Space Scores for nominees from Segal and Cover scores: Common Space Score = 0.4401225 - 0.9148797 (Segal and Cover).

We then applied that transformation to all nominees to derive the Common Space Scores. The (euclidean) distance variable was calculated according to the following formula: Euclidean Distance = (cs1 - cs_nom) (where cs1 is the first dimension Common Space Score for Senators, and cs_nom is the Common Space Score for Nominees).

75. Cameron et al., supra note 15, at 530-31.
tures the idea that some Presidents are simply in a better position to attain approval of their nominees than are others. We attend to this idea as did Cameron and his colleagues, with a variable that takes on the value of one if the President’s political party controls the Senate and the President is not in the fourth year of his term of service. Otherwise, the value is zero. The second control variable, which we include for all the obvious reasons, indicates whether the President and the individual Senator are of the same political party. If they are, the variable is coded one. If they are not, the variable is coded zero.

E. Summary of the Model

To summarize, we argue that four factors explain the votes of individual Senators on nominees to the U.S. Supreme Court.

1. Lack of Qualifications: The lower a nominee’s qualifications (see Figure 3), the lower the probability of a Senator voting in favor of the nominee.

2. Ideological Distance: The further the ideological distance between a nominee (see Figure 4) and a Senator, the lower the probability of a Senator voting in favor of the nominee.

3. Strong President: If a President is strong, the higher the probability of a Senator voting in favor of the nominee.

4. Same Party: If a President and Senator are of the same political party, the higher the probability of a Senator voting in favor of the nominee.

Given the way we code the variables, we expect the estimated coefficients on the first two variables to be negatively signed, and on the second two variables, positive.

Finally, we incorporate into our model a crucial fifth variable—one representing the interaction between Qualifications and Ideological Distance. This enables us to determine the extent to which Senators vote against unqualified nominees, who are ideologically distant from them (as we might hypothesize), apart from any independent effects the two variables may exert on that vote. For example, just as Republicans found President Clinton’s ethical failings to be far more serious than did Democrats, we might expect liberals to be far more affected by the charges against Clarence Thomas, and conservatives far more troubled by the ethical allegations against Abe Fortas. Whether this “conditional” response to a nominee’s professional integrity (or, more pointedly, lack thereof) is the result of motivated reasoning, the psy-

76. Id. at 529-30.
77. See supra p. 1160 fig.3.
78. See supra p. 1162 fig.4.
chological reflex that causes humans to believe those arguments they wish to believe, or sheer hypocrisy is beyond the scope of this study; we simply hypothesize that the data will indicate behavior consistent with the response.

IV. RESULTS OF THE ANALYSES

Because the outcome of our dependent variable, a Senator’s vote, is binary (yea or nay), standard OLS regression is inappropriate. Accordingly, we explore the effect of the variables on the vote choice via a probit model, which we estimate using maximum likelihood.

Table 2 summarizes the results, which, all in all, appear satisfactory. Each coefficient runs in the right direction; each is statistically significant at \( p \leq .01 \); and none is trivial in size. We thus could say much the same of our exercise as did the Cameron team about theirs: “Judged by an array of statistical criteria, the model was very successful.”

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>COEFFICIENT</th>
<th>(STANDARD ERROR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of Qualifications</td>
<td>-0.678**</td>
<td>(0.239)</td>
</tr>
<tr>
<td>Ideological Distance</td>
<td>-0.641**</td>
<td>(0.249)</td>
</tr>
<tr>
<td>Lack of Qualifications × Ideological Distance</td>
<td>-9.009**</td>
<td>(0.920)</td>
</tr>
<tr>
<td>Strong President</td>
<td>0.900**</td>
<td>(0.103)</td>
</tr>
<tr>
<td>Same Party</td>
<td>0.367**</td>
<td>(0.089)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.507**</td>
<td>(0.094)</td>
</tr>
</tbody>
</table>

| N                                      | 2461        |
| Log-likelihood                         | -584.399    |
| \( \chi^2 \)                            | 942.23      |

These basic results now noted, let us consider how they inform the major concern of this Essay: the relative roles of politics and professional merit in the confirmation of Supreme Court Justices.\(^{81}\) Beginning with politics, Table 2 tells us that the coefficient produced by the Ideological Distance variable is negative and significant, indicating that, in fact, politics does play a role in confirmation proceedings: As the ideological distance between a Senator and nominee increases, the probability of a nay vote increases. So

---

80. ** indicates \( p \leq .01 \).
81. Note that Table 2 displays probit coefficients, which are not as easy to interpret as, say, OLS regression coefficients. Accordingly, in our discussion of the results here and elsewhere, we transform them into probabilities.
too, as we can see in the left panel of Figure 6, infra, this statistical finding is not, especially if we keep in mind that nearly eighty-five percent of all votes are “yays,” without substantive import. Figure 6, infra, demonstrates that the probability, across the twenty-six nominees under investigation, of a favorable vote varies depending on a Senator’s proximity to the candidate under consideration. When we set all other variables at their mean (or median [in the case of variables that take on values of 0 or 1], except for the interaction term), the likelihood, on average, of a Senator voting for a candidate is .235 when that Senator and the candidate are ideological extremes (the black line). That figure increases by a factor of 4.23, to .994, when they are at the closest levels (the dashed line).

**Figure 6**

**The Effect of Ideology and Qualifications on the Votes of Senators over Supreme Court Nominees from Earl Warren to Stephen Breyer.** Each curve represents the probability density of voting yea on the nominee, accounting for the uncertainty in our estimates. All variables are set at their sample means (or medians) (for each panel, the variable on the x-axis is interacted with the sample mean of the other variable). The left panel shows these probabilities for ideologically distant nominees (black) and ideologically proximate nominees (dashed); the right panel, for highly unqualified nominees (black) and highly qualified nominees (dashed).

---

Nearly as important as the candidate’s policy preferences (relative to the Senator’s), though, is his or her professional merit. To see this, consider the right panel which demonstrates the probability of a Senator voting for a nominee on the basis of the nominee’s qualifications. Notice that when a nominee is perceived as highly unqualified (the black line) and all other variables are at their mean (or median, except for the interaction term), the likelihood of a Senator casting a nay vote is $(1 - .75 = .25)$; that probability decreases 125-fold to $(1 - .998 = .002)$ when the nominee is highly qualified (the dashed line).

Figure 6 depicts the likelihood of a yea vote when all the variables (except those of interest) are set at their mean (or median, except for the interaction term). Also worthy of exploration are other possible scenarios, such as the one we display in Figure 7, infra: when the President is at his lowest level of influence because the Senator and the President hail from different parties, the President’s party does not control the Senate, or he is in the fourth year of his term of service (and all other variables are set at their mean or median, except for the interaction term). Note, first, the effect of ideological distance (depicted in the left panel): if the ideological distance is minimal (the solid line), the Senator will still, in all likelihood, vote for the President’s nominee (.914 probability) even though the President is quite weak. But the odds turn against the President (.025) when the Senator and the nominee are ideological extremes (the dashed line). Now consider the near-parallel effect of qualifications (shown in the right panel): Presidents—even those with severely limited political clout—will have a far easier time attaining the confirmation of their candidate if Senators perceive that candidate as highly qualified (.915 for the solid line) than if they do not (.161 for the dashed line).
**Figure 7**

The effect of ideology and qualifications on the votes of Senators over Supreme Court nominees, from Earl Warren to Stephen Breyer, when the President is at his lowest level of influence. Each curve represents the probability density of voting YEA on the nominee, accounting for the uncertainty in our estimates. The left panel shows the probability of a Senator casting a YEA vote when the Senator and the nominee are ideologically proximate (the black line) and ideologically distant (the dashed line).

Taken collectively, these results indicate that both ideology and qualifications have a significant, independent effect on the Senate’s decision to confirm. However, it is the interaction between the two that provides the greatest explanatory power. Senators will most certainly vote for candidates who are ideologically close and well qualified, and they also will almost certainly vote against candidates who are distant and not qualified. And, yet, while the odds are high that they will vote for an undeserving candidate who is ideologically proximate (for example, Southern Democrats and Clement Haynsworth)—thereby underscoring the role of politics—it is also the case that they will, under certain conditions, support a distant candidate if they perceive that candidate to be highly meritorious (for example, Republicans and Ruth Bader Ginsburg), thereby underscoring the role of qualifications.

Figure 8, *infra*, explores this relationship between ideology and qualifications. The left panel compares qualified candidates (the top vertical lines, representing a ninety-five percent confidence interval) and unqualified candidates. The right panel shows the probability of a Senator casting a YEA vote when a nominee is perceived as highly qualified (the black line) and highly unqualified (the dashed line). In both panels we set the President at his weakest: he and the Senator are from different parties and his party does not control the Senate or he is in the last year of his term. All other variables are set at their sample means (or medians) (for each panel, the variable on the x-axis is interacted with the sample mean of the other variable). We generated these figures using Zelig. See King et al., *supra* note 82.

---

83. The right panel shows the probability of a Senator casting a YEA vote when a nominee is perceived as highly qualified (the black line) and highly unqualified (the dashed line). In both panels we set the President at his weakest: he and the Senator are from different parties and his party does not control the Senate or he is in the last year of his term. All other variables are set at their sample means (or medians) (for each panel, the variable on the x-axis is interacted with the sample mean of the other variable). We generated these figures using Zelig. See King et al., *supra* note 82.
candidates (the lower vertical lines, also a ninety-five percent confidence interval) over the range of the Ideological-Distance variable when the President is weak and all other variables are at their mean (except for the interaction term).84 The right panel is a comparison between ideologically close candidates (the top vertical lines) and those that are distant (the lower vertical lines), over the range of the Lack-of-Qualifications variable.85 The results are intriguing and underscore the claim that a significant difference exists, as indicated by the nonoverlapping vertical lines, between qualified candidates and those less qualified in terms of their likelihood of attaining a position on the Court, even at the lowest levels of ideological distance. Candidates who are highly meritorious and ideologically proximate are virtually assured a position on the Court, but the probability of confirmation uniformly decreases as the Lack-of-Qualifications variable moves from the lowest (most qualified) to highest (least qualified) levels.

FIGURE 8

The relative effects of ideology and qualifications on the votes of Senators over Supreme Court nominees, from Earl Warren to Stephen Breyer.86 The left panel (Ideological Distance) shows the probability of a Senator casting a Yea vote when the nominee is highly qualified (darker vertical lines) and not qualified (lighter vertical lines) over the range of the Ideological Distance variable. The right panel (Lack of Qualifications) shows the probability of a Senator casting a Yea vote when the Senator and the nominee are ideologically proximate (darker vertical lines) and ideologically distant (lighter vertical lines) over the range of the Lack of Qualifications variable.

84. The pooled first difference mean (standard deviation) = -0.8073 (0.158).
85. The pooled first difference mean (standard deviation) = -0.7477 (0.1741).
86. In both instances the President is “weak” and all other variables are set at their mean (or median, except for the interaction term). We generated these figures using the program Zelig. See supra note 82.
V. DISCUSSION

Virtually all contemporary writing on the confirmation of Supreme Court nominees has it exactly right: politics plays a critical role. Our statistical modeling exercise leaves little doubt that Senators are more likely to vote for nominees who share their policy preferences; they also are more likely to support the candidates of Presidents who share their partisanship.87

But that same modeling exercise also leaves little doubt about the crucial role of qualifications. Whether Senators perceive a candidate as meritorious affects their votes and, indeed, exerts an effect nearly as strong as ideological proximity. Hence, our results give some empirical teeth to Watson and Stookey’s assertion that the current “justices are [not] less well qualified on some objective measure . . . than justices of the past.”88

Of course, this is not to say that the President always taps the “most capable,” the “best qualified,” or the “most meritorious” person at any given time to fill any given vacancy. It is possible that there was someone in 1969 more qualified than Warren Burger to serve as Chief Justice or more meritorious than was Stephen Breyer in 1994—to name just two of the twenty-six nominees we examined. It is necessary, however, that any proposals to reform the process account for the chief lesson of this study: qualifications do not appear to play a trivial role in the confirmation of Justices. This result, at the very least, should be a cause for pause before we offer (and policymakers consider) schemes for abrupt change, such as a tournament among judges, which may not only perpetuate perilous norms,89 but also, by virtue of eliminating all but federal judges, eliminate the “most capable,” the “best qualified,” and the “most meritorious” person from consideration for a seat on the nation’s highest court.90

87. See supra p. 1168 tbl.2.
88. Watson & Stookey, supra note 8, at 222.
89. See Lee Epstein et al., The Norm of Prior Judicial Experience and Its Consequences for Career Diversity on the U.S. Supreme Court, 91 CAL. L. REV. 903 (2003).
90. This is not the place to conduct a detailed investigation of whether nominees with or without prior federal judicial experience make better Supreme Court Justices. (For this type of analysis, see Workshop on Empirical Research in the Law, On Tournaments for Appointing Great Justices to the U.S. Supreme Court, 78 S. CAL. L. REV. 157 (2004).) However, it is worth noting that in a 1998 survey (the most recent we could locate) asking scholars to rate the fifty-two Justices appointed in the twentieth century (from Holmes through Breyer), only one appellate court judge (Harlan II) received an “excellent” rating, even though there were plenty of potential candidates. Instead, the balance of the list consisted of the familiar “greats”: Holmes and Brandeis (tied for the first and second spots), Cardozo, Brennan, Warren, Hughes and Black, Stone, Frankfurter, and R. Jackson. Michael Comiskey, Has the Modern Senate Confirmation Process Affected the Quality of U.S. Supreme Court Justices? (1998) (unpublished manuscript), cited in Abraham, supra note 11, at 372. Interestingly enough, as Professor Abraham notes, Comiskey answers his research question in the negative.
Can we measure the quality of a judge’s past performance? Should the President weigh this factor in making Supreme Court appointments? I will argue that the answer to both questions is the same: “Yes, but with significant qualifications.” Yes, we do have some indices for assessing judicial performance—but those measures are rough and incomplete. Yes, the President should consider past judicial performance and its objective indicators—but only to a limited degree.

These comments are prompted by two important recent articles by Stephen Choi and Mitu Gulati. They make a more ambitious claim: that we can measure judicial quality well enough to make it the predominant element in Supreme Court appointments. In their initial article, they proposed a tournament, an objective contest among federal appellate judges in which the prize is a Supreme Court appointment. In their follow-up article, they run an initial tournament using recent data about federal appellate decisions. They hedge their conclusions about “who is the fairest in the land,” but the clear
import seems to be that Richard Posner should be the next Supreme Court appointment.5

While the tournament idea is intriguing, it is far too ambitious. In my view, a conscientious President probably would not choose Justices on the basis of such a tournament, partly because we cannot measure professional merit well enough and partly because professional merit is not the only factor that the President should find relevant. Nevertheless, Choi and Gulati have made a valuable contribution by showing how objective measures of quality can improve the appointments process. Simply making the tournament data available on a regular basis through a reliable information source could help temper partisan claims and provide guidance for the public. Moreover, the data in their follow-up article raises tantalizing questions about the appellate process and, in particular, about the dynamics of judicial influence. Those questions would be well worth pursuing even if the results had nothing to tell us about how to appoint Justices.

II. INDICATORS OF JUDICIAL QUALITY

Supreme Court Justices engage in four major tasks: (1) they vote on which cases to accept, (2) they vote on the outcomes of cases, (3) they write majority opinions, and (4) they write dissents and concurrences.6 By contrast, federal appellate judges only engage in two of these tasks.7 Federal appellate judges rarely write separate opinions, so the fourth factor is less significant in assessing their performance. Also, nearly all appeals at the circuit level are mandatory, which makes the first factor less significant. This leaves two major tasks for federal circuit judges: voting on the merits and writing opinions. How well a judge performs these tasks on a lower court presumably has some predictive value about potential performance on the Supreme Court.8 Choi and Gulati attempt to measure circuit judges’ performance of these tasks along three dimensions: productivity, quality of opinion writing, and independence.9 Of these, as we will see, the at-

---

5. Id. at 74, 113 tbl.H. I know a number of academics who agree—including some who are liberals or skeptics about Posner’s specialty, law and economics—but I do not know of anyone who thinks the prospect is likely.

6. Justices also engage in other tasks, such as considering stay orders or approving changes in federal procedural rules, but these seem less central to the role.

7. The current norm is for Presidents to nominate circuit judges to the Supreme Court. Lee Epstein et al., The Norm of Prior Judicial Experience and Its Consequences for Career Diversity on the U.S. Supreme Court, 91 CAL. L. REV. 903 (2003); see also Choi & Gulati, Empirical Ranking, supra note 2, at 40. Whether this norm is desirable is outside the scope of this Essay.

8. How much predictive value is another question, which we will consider in Part IV, infra.

9. See Choi & Gulati, Empirical Ranking, supra note 2, at 33, 42.
tempt to measure opinion quality via citation rates is probably the most successful.

A. Productivity

One of the stated purposes of the tournament is to encourage judges to work harder as they compete for higher rankings and Supreme Court nominations. One of the indices of judicial performance, consequently, is productivity. Since circuit judges have discretion over which of their opinions are published, Choi and Gulati argue that judges with a higher number of published opinions are investing more time and effort in polishing opinions for publication. This is quite plausible, but there are alternative explanations for high publication rates that make these rates a weak indicator of effort.

One possible explanation for a high publication rate could be that a judge has trouble identifying which cases are important enough to serve as precedents and therefore deserve published opinions. Such an inability would be a handicap in a Supreme Court Justice. One of the Justice’s tasks is to decide which cases deserve Supreme Court review; thus a judge who cannot accurately gauge the significance of a case is at a disadvantage.

Another possible explanation is self-centeredness. Some judges may consider all of their output to be worthy of preservation for the ages, simply because it is theirs. A related explanation is that a judge may fail to appreciate the costs that publication imposes on others. The two other members of the panel must invest additional time in reviewing and commenting on a published opinion, and additional published opinions increase the size of the pool of authority that lawyers must search. Productivity based on self-centeredness does not speak well of judicial quality or recommend promotion to the Supreme Court.

An alternative explanation for high publication rates is consistent with the idea that some judges invest more effort in perfecting their opinions for use as precedents. Judges may vary in how important they consider this function of opinions, and some judges may focus more heavily on their role as generators of precedents at the expense of their dispute settlement role. A published appellate opinion not only provides a precedent, it resolves a specific piece of litigation. It is important for the functioning of the legal system as a whole that appeals be correctly decided, which means not only that the correct rules of law are announced for the future but that the rules are relevant and correctly applied to the specific case. This may take tedious

10. Choi & Gulati, Tournament, supra note 2, at 304.
investigation of the trial record. A judge who is willing to cut corners on fidelity to the record can publish more opinions, and those opinions may be well regarded as precedent. Yet, that judge also decreases the ability of the legal system to respond accurately to the facts of cases and degrades the overall performance of the system (besides being unfair to individual litigants). Furthermore, individual cases themselves sometimes have substantial social importance (think of the Microsoft antitrust litigation or of redistricting challenges), so it is important for judges to invest time in scrutinizing the record even if doing so does not increase the opinion’s precedential value.

A final, and even simpler, explanation is that some judges need not invest a high level of effort in order to get a high publication count, either because they themselves are facile writers or because they are able to recruit and motivate highly productive law clerks. Being able to write easily, or successfully motivate clerks to do so, is a valuable trait. It is not, however, a gauge of effort. Moreover, judges are probably either good writers when they are appointed (or good supervisors) or they are not; either way, providing incentives is unlikely to make much difference.

The bottom line is that productivity may measure good judicial traits such as effort or writing ability or bad judicial traits such as self-centeredness or sloppy treatment of facts. Without knowing more, we cannot be sure of what we are measuring or whether it is something we want to encourage.

B. Citations as a Measure of Opinion Quality

Are citation counts, of any kind, a gauge of quality? In the context of appellate judges, it seems plausible to think so. For citations to be unrelated to quality, we would have to assume that judges are just as likely to cite and rely on muddled or ill-reasoned opinions as on cogent and logical ones. This would indicate a serious problem with the functioning of the federal judiciary. Presumably, judges tend to cite cases which best support their arguments or most require distinguishing or rebuttal. Consequently, they should favor strong opinions over weak ones in their citation practice.

Yet we should not expect citations to be a very precise measure of quality. Too many other factors may influence how often a judge’s opinions are cited: the “luck of the draw” (how many significant cases are drawn by panels on which that judge sits), the author’s reputation (which may imperfectly reflect actual quality), the quality of the judge’s law clerks in a particular year, ideological or personal affiliations (or antipathies) between judges, and just plain luck (perhaps an opinion gets cited more often because its name is short and memora-
ble or because the facts are striking).\footnote{Also, the citation figures do not seem to measure quality for members of the D.C. Circuit, who tend to get fewer citations because their workload is specialized, centering on administrative law. See Choi & Gulati, \textit{Empirical Ranking}, supra note 2, at 40 n.30 (recognizing that the different nature of the D.C. Circuit’s docket from that of the other circuits might make total citations a less meaningful comparison, but including D.C. Circuit judges in the tournament because other citation figures, such as citations to top twenty opinions and citations by law review, might be more meaningful).} For these and other reasons, it seems right to say that “[c]itations are at best a crude and rough proxy for measuring influence.”\footnote{William M. Landes et al., \textit{Judicial Influence: A Citation Analysis of Federal Courts of Appeals Judges}, 27 J. LEGAL STUD. 271, 271 (1998).} In turn, influence may only imperfectly reflect quality.

Moreover, we may not want circuit judges to strive to maximize their citation rates. In seeking to increase their influence, judges may reach for broader holdings by ignoring the factual nuances of specific cases. Or in trying to write influential opinions, they may simply not pay much attention to the litigation record. They might also be likely to insert unnecessary dicta or to address issues not raised by the lawyers or addressed by the trial judge. None of these seem to be particularly beneficial behaviors.

Although these points suggest the need for caution in relying on citation counts, it remains true that citation counts tell us something important about how useful opinions are to other judges. A judge whose opinions are consistently useful to others is probably doing something right, while a judge whose opinions are rarely cited is probably performing badly.

\textbf{C. Independence}

“Knee-jerk” is not usually considered a good adjective as applied to judges. We presumably do not want judges who always vote for the most liberal or conservative result without considering specific legal arguments. It is probably even worse to have judges who always vote with members of the same political party without regard to the merits of cases. Thus, Choi and Gulati are right to include independence as one of their measures, but their measurement of this characteristic is problematic.

Choi and Gulati measure independence by counting disagreements with other judges appointed by Presidents of the same political party.\footnote{See Choi & Gulati, \textit{Empirical Ranking}, supra note 2, at 63.} This does not seem to be a particularly apt measure. A Republican-appointed judge who often votes on the opposite side from others may be more moderate, but equally well could be an ideological extremist who is parting company with less ideological colleagues. Moreover, since we are talking about relatively small num-
bers of judges in each circuit, it is hard to control for other variables. Some judges may have been appointed by members of the same party but have sharp personality clashes with each other, or one may be markedly more competent than the other. This could lead to frequent disagreements but would not tell us much about whether a judge is voting on a purely ideological basis.

More fundamentally, independence is not a neutral concept. What one of us sees as an ideologue, another may see as a person of principle. Judges who vote on the basis of expediency rather than principle may disagree frequently with their more principled colleagues. Or to put it another way, one might view “independence” as a euphemism for “inconsistency” and hence as an undesirable trait.

Deciding how much weight to give to independence or how to measure it requires a normative judgment. Formalists and pragmatists may have different concepts of what kind of independence is required by judges. For the formalist, it is the ability to strictly follow a set of rules, regardless of pressure to reach a more congenial result. For the pragmatist, it is the ability to be open to opposing arguments, consider the implications of a decision, and reach a balanced conclusion. These translate into very different conceptions of judicial independence.

III. PROBING THE MEASURES OF OPINION QUALITY

Of the three basic criteria discussed in Part II—productivity, opinion quality, and independence—it is the second factor that seems most useful in assessing judicial performance. As we have seen, productivity is an ambiguous quality, which might reflect either well or badly on judicial performance. Independence is normatively charged and hard to measure. Thus, the remainder of this Essay will focus primarily on opinion quality as the measure of judicial performance. Obviously, other highly desirable traits may not be captured by this measure.

Opinion quality is assessed through citation counts. Although citation counts do seem related to the quality of judicial opinions, various different citation counts can be used and the data can be handled in different ways. We need to know more about the processes that generate this data before we can decide on the proper method of measurement. Choi and Gulati have assembled some extremely interesting data, which provide a basis for a more detailed analysis.

This analysis is significant for reasons that go beyond the tournament idea. Citation practices—which opinions are cited, by whom, and for what purpose—are a basic part of the legal system. But we

14. See id. at 48-49.
know little about the processes involved. The data presented by Choi and Gulati help shed some light on these broader questions.

A. Power Laws and Normal Distributions

In trying to understand the process that produces citation rates, one of the most important clues is the shape of the statistical distribution. Different kinds of processes characteristically produce different sorts of distributions. For purposes of this Essay, we are primarily interested in two types: the normal distribution and the power-law distribution.

The best-known statistical distribution is the “bell curve,” or normal distribution, which is often associated with the idea of random variation.\textsuperscript{15} Bell curves are associated with a certain type of random process. According to a basic statistical law, “the sum of a large number of independent random variables will be approximately normally distributed almost regardless of their individual distributions.”\textsuperscript{16} Thus, when many small factors combine to produce a result and each factor has an element of randomness, the result is likely to be a normal distribution.

In assessing how close a distribution comes to being normal, there are three useful parameters. The first involves the center of the distribution. For the normal curve, the median, the mean (what most people mean by the “average”), and the mode are identical. The second parameter is skew, which measures asymmetry. A normal curve is symmetrical rather than being stretched in one direction. The skew parameter is zero for the normal distribution. A high skew means that a distribution is bunched on one side and stretched out on the other. The third parameter is called kurtosis. Kurtosis measures whether a curve is flattened out or unusually peaked, compared with the normal distribution. The normal distribution has a kurtosis of three.\textsuperscript{17}

Apart from the normal curve, we are also interested in another type of distribution—the power-law distribution. Complex, nonlinear systems have a characteristic distribution of outcomes: a “high frequency of small fluctuations, punctuated by the occasional large shift[] in system conditions.”\textsuperscript{18} Rather than following the familiar normal distribution, outcomes in complex systems often follow what

\textsuperscript{16} Id. at 109.
\textsuperscript{17} See id. at 63-65.
are called power laws—that is, the frequency of an event is often given by its magnitude taken to a fixed negative exponent.

Some additional explanation of power laws may be helpful for the non-mathematically inclined. Albert-László Barabási, a physicist who studies complex networks, explains how power laws work. Contrasting power laws with the normal curve governing characteristics such as human heights, he points out that a frequency distribution “following a power law is a continuously decreasing curve, implying that many small events coexist with a few large events.” For example, “[i]f the heights of an imaginary planet’s inhabitants followed a power law distribution, most creatures would be really short. But nobody would be surprised to see occasionally a hundred-feet-tall monster walking down the street.” Such “outliers” are much less likely when a normal distribution is involved. Or, in more technical terms, “[b]ell curves have an exponentially decaying tail, which is a much faster decrease than that displayed by a power law.”

Caselaw can be considered a network of cases linked by citations. Power laws seem characteristic of complicated networks. Some examples include the World Wide Web, where the number of links to a particular site follows a power law; the number of citations to a given physics paper; and even the number of other actors with whom a given Hollywood star has appeared. On the Web, for example, about ninety percent of a sample of two hundred million web pages are the targets of ten or less links, while about three pages had roughly a million other pages linking to them. Similarly, students of bibliometrics refer to Lotka’s law, under which productivity follows an inverse-square relationship: if one hundred scholars in a group each publish one paper annually, then twenty-five will publish two, eleven will publish three, six will publish four, and so forth, with a single scholar producing ten papers. In other words, the number of scholars who produce $N$ articles per year is proportionate to $1/N^2$. As it turns out, Supreme Court opinions also appear to follow a power law in terms of their frequency of citation.

---


22. Id.

23. Id. at 68 n.1.

24. Id. at 67-69.

25. Id. at 57-58.


Several possible mechanisms might result in a power-law distribution for judicial citations. One is the superstar phenomenon.\textsuperscript{28} Suppose that judges differ in quality. The top judge’s opinions are slightly better than everyone else’s. So when that judge has written on an issue, her opinion is the one cited by other judges. When she has not written on an issue but the second-best judge has done so, the second-best judge’s opinion gets cited instead; and so forth. Small differences in quality can translate into large differences in rankings—just as small differences in GPAs can translate into large differences in class ranks.

We can get some sense of how the superstar phenomenon works with a simple model. Assume that the chances that a particular judge will get to write about a particular issue is $1/a$ (in other words, this is the probability that the judge will sit on the first case in the circuit raising that issue and be assigned to write the opinion). Thus, $1/a$ is the proportion of issues that any particular judge can write precedential opinions about. In this simple model, once an issue has been decided in a given circuit, later opinions cite that as the dispositive opinion in the circuit because it binds future panels; later judges in the circuits have no reason to do any more than cite that case. Then the best judge will write on $1/a$ of the issues, and those opinions will be the exclusive citations when those issues arise, so that judge receives a $1/a$ share of the cites. The second-best judge has a $1/a$ share of the remaining cites, or $(1/a)(1 - 1/a)$. By the time we get to the “$(n + 1)$ judge,” the number of remaining cases where no previous judge has written is $(1 - 1/a)^n$, and so that judge gets cited in $(1/a)(1 - 1/a)^n$ of the cases. Note that the ratio between the “$n$ judge” and the “$(n + 1)$ judge” in terms of citations is constant: $(1 - 1/a)$. This is a power-law distribution.\textsuperscript{29}

Note that in this model, the ranking of citations perfectly replicates the ranking of quality, but the scale is distorted, since even a tiny difference in opinion quality means that the top judge will be cited in preference to the other judges in every case where the top judge has written.\textsuperscript{30} The superstar effect can also appear even if there

\textsuperscript{28} See Choi & Gulati, Empirical Ranking, supra note 2, at 72.

\textsuperscript{29} A fuller model would introduce some complications. Certain judges may sit on circuits where more issues of first impression arise, which would increase their chances of hearing such cases. Similarly, by reason of seniority, some judges may be more likely to get assigned the interesting opinions. Also, if the number of novel issues is not proportionate to caseload, judges from smaller circuits might have an advantage because fewer judges would be splitting up the interesting cases. Conceivably, influential judges might sometimes be cited when they merely follow existing circuit precedent, rather than making new law. Finally, it probably is not true that a higher-ranking judge is always cited in preference to a lower-ranking one, so we might need to add some random variation.

\textsuperscript{30} For a summary of various situations that can produce a superstar distribution, see Mitu Gulati & Veronica Sanchez, Giants in a World of Pygmies? Testing the Superstar Hypothesis with Judicial Opinions in Casebooks, 87 IOWA L. REV. 1141, 1181-89 (2002).
are no differences in quality between judges. If consumers tend to cite cases in proportion to their existing popularity, with only a small chance of citing something that has not been cited previously, the distribution of product types will tend to something called the Yule distribution, which is close to being an inverse-square power distribution.31

Other processes might produce a normal distribution rather than a power-law distribution. The citation count for a given opinion might depend on a whole series of factors, each of which varies randomly. For example, the list of factors might include:

1. The judge’s past experience with issues in the field.
2. The judge’s IQ.
3. How much time the judge can devote to the particular opinion, given other responsibilities.
4. The quality of the lawyers in the case (it is presumably easier to write a good opinion if the briefs are of high quality and the case has been well litigated).
5. Who else is on the panel and how many changes they demand.
6. Whether the opinion is well classified by the digest systems, and whether the judge used the right key words that will show up in a computer search.
7. Whether the opinion happens to be cited in the next few cases on the issue.

If we assume that each of these and similar factors has an element of randomness, their combined effect is likely to look like a normal distribution. On the other hand, a power-law distribution suggests that some kind of feedback is at work. With these observations in mind, we consider the distributions of various measures of opinion quality.

B. Invocations

We begin with invocations, that is, with the number of times the judge is identified by name in later opinions as the author of a major-

---

31. See Kee H. Chung & Raymond A.K. Cox, A Stochastic Model of Superstardom: An Application of the Yule Distribution, 76 REV. ECON. & STAT. 771 (1994). The Yule distribution can be approximated by $1/[N(N + 1)]$, see id. at 773, which is bounded above by (and converges with) $1/N^2$. Thus the statement in the text that the Yule distribution is “close to” an inverse-square relationship. In another model, reputation effects and search costs can produce the superstar effect even if there are no quality differences. See Moshe Adler, Stardom and Talent, 75 AM. ECON. REV. 208 (1985).
As we will see, this distribution looks nothing at all like a bell curve. In fact, it is more extreme than one would expect from a power-law distribution.

Table 1 shows that something odd is going on with invocation rates.

<table>
<thead>
<tr>
<th>TOTAL INVOCATIONS</th>
<th>NUMBER OF JUDGES</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-9</td>
<td>88</td>
</tr>
<tr>
<td>10-24</td>
<td>8</td>
</tr>
<tr>
<td>25-49</td>
<td>0</td>
</tr>
<tr>
<td>50-74</td>
<td>0</td>
</tr>
<tr>
<td>75-99</td>
<td>0</td>
</tr>
<tr>
<td>100-124</td>
<td>1</td>
</tr>
<tr>
<td>125-149</td>
<td>0</td>
</tr>
<tr>
<td>150-174</td>
<td>0</td>
</tr>
<tr>
<td>175-199</td>
<td>1</td>
</tr>
</tbody>
</table>

(Note that what would otherwise be the first row here (0-24) is split into two, in order to give a better sense of the shape of the distribution.) The distribution consists of two outliers (who are actually Posner and Easterbrook) along with a large clump. The summary statistics confirm the oddness of this distribution: the kurtosis is a remarkably high 54.7 (recall that the kurtosis of a normal curve is three).

For purposes of comparison, imagine that the same table showed someone’s journal of food expenses. On most days the person lives frugally, spending less than $10 a day; on other days the person lives more expansively, spending up to $25. But on two days out of a hundred, the person’s food expenses leap over $150. The obvious inference is that something special happened on those days: either the person splurged at a really fancy restaurant or invited a group of people to dinner. These were obviously not just days when the person

33. Table 1 was constructed by collating figures from Table F of Choi & Gulati, *Empirical Ranking*, supra note 2, at 104-07.
34. Posner had 176 invocations and Easterbrook had 103 during the time span studied by Choi and Gulati. *Id.*
35. *Id.* at 107.
felt a little hungrier and ordered extra food. Similarly, something special seems to be happening with the invocation rates for Posner and Easterbrook.

One possibility is that Posner and Easterbrook are simply judges of an entirely different caliber than their fellows, like two NBA players who have wandered into a lawyers’ after-work basketball game. While they may well be exceptionally good judges, however, it seems unlikely that differences in ability level can account for the magnitude of the difference in invocation rates. The third-highest-ranking judge is Guido Calabresi, with twenty-three invocations.\textsuperscript{36} Posner and Easterbrook are undoubtedly very bright, but are they five times as bright as Calabresi, not to mention the other eminent judges on the list?

Another possibility is the superstar phenomenon—judges invoke only the very best names, given a choice, so a judge who is just a little better than his fellows will get disproportionately more invocations. Without a good model of how reputations develop, it is hard to test this explanation. We can get some sense of its plausibility by assuming that a judge’s prestige tracks the number of legal issues on which a judge has written the leading opinion. Under the basic superstar model, this is simply equal to the number of cases on which a judge is cited, since courts are presumed to cite only the leading (“best”) opinion. But the distribution of invocations does not fit with this explanation. The problem is that no matter how good Easterbrook and Posner are, they first have to sit on a case where an issue is presented in order to write the leading opinions on that issue. There must be a great many legal issues that, for one reason or another, have not been the subjects of opinions by Easterbrook or Posner: either they did not happen to be on a panel on which these issues were raised or someone else wrote the opinion. In these other cases, the superstar phenomenon should lead to large numbers of citations of leading opinions by the third- or fourth-best judges on issues that never reached Posner and Easterbrook. Hence these second-tier judges should have written almost as many leading opinions as Posner and Easterbrook, and their invocation rates should not be drastically lower. But this is not what we see in the data.

Recall that, in the basic model, the superstar effect produces a constant ratio between citations as we move one rank in the ratings. But for invocations, the ratio between judges 1 and 2 is 103/176, or 0.58; the ratio between judges 2 and 3 is 23/103, or 0.22; and the ra-

\textsuperscript{36} Id. at 104 tbl.F.
The ratio between judges 3 and 4 is 19/23, or 0.82.\textsuperscript{37} Moreover, the first two ratios are hard to account for on the basis of the superstar effect. Consider the 0.58 ratio between the first and second judges. In the basic superstar model,\textsuperscript{38} the ratio between judges who adjoin in the rank order is \((1 - 1/a)\), and if we set this equal to 0.58, we get \(1/a = 0.42\). (Remember that \(1/a\) is the likelihood that a judge will write the controlling circuit opinion on the issue.) But this seems implausibly high—it would require that the top judge happen to sit on the panel (and get assigned the opinion) in 42\% of the cases raising issues of first impression in the circuit.\textsuperscript{39} For similar reasons, the 0.22 ratio between judges 2 and 3 is even harder to square with the superstar model.\textsuperscript{40}

Invocations reflect reputations rather than raw citation figures, but it is hard to see why the falloff between reputations should be so steep as we move from the judge who is considered to be the best in the country to the second-best and then to the third-best judge. We cannot exclude this possibility, however, without a better theory of how judicial quality is translated into reputations.

Another plausible possibility is that Easterbrook’s and Posner’s names are invoked for reasons other than general judicial reputation. Perhaps they are invoked in cases involving their special area of expertise: economic analysis.\textsuperscript{41} It would be entirely understandable that these two judges should be the dominant judicial figures on economics issues.

\textsuperscript{37} The numbers used to calculate these ratios are the total invocations to the top four judges—Posner, Easterbrook, Calabresi, and Wilkinson, respectively—as reported by Choi and Gulati. \textit{Id.}

\textsuperscript{38} See supra text accompanying notes 28-30.

\textsuperscript{39} Consider a court comprised of \(N\) members. What are the odds that a given judge will not be on a panel? When we pick the first judge on the panel, the odds are \((N - 1)/N\) of getting someone other than the specified judge. Now there are only \((N - 1)\) judges left in the pool, so the odds of missing the specified judge the second time are \((N - 2)/(N - 1)\). Similarly, the odds of missing the specified judge a third time are \((N - 3)/(N - 2)\). If we multiply these together and simplify, everything cancels out except \((N - 3)/N\), so the chances of missing the specified judge all three times is \((1 - 3/N)\). Hence, the judge will be on \(3/N\) of the panels. But Posner is from the Seventh Circuit, and we know that there were ten judges from that circuit in the tournament, see Choi & Gulati, Empirical Ranking, supra note 2, at 41 tbl.1, and this does not include post-1998 appointments, senior judges, or judges sitting by designation, all of whom serve on panels. Even for \(N = 10\), we would get only a 30\% chance of serving on a particular panel, so it is unlikely that Posner would have the opportunity to sit on panels hearing 42\% of the cases that raise issues of first impression.

\textsuperscript{40} With \((1 - 1/a)\) equal to 0.22, \(1/a = 0.78\). Following up the reasoning of the previous footnote, this would require that the judge sit on 78\% of the panels hearing issues of first impression. Even in a smaller circuit like the Seventh, any one judge sits on less than half this percentage of cases.

\textsuperscript{41} Calabresi is also a major figure in the law and economics movement, but his major work in this genre came earlier in his career and was focused on the specific field of torts.
Economics issues are important in some Supreme Court cases, so Posner’s and Easterbrook’s expertise should count in their favor. But these cases may not be the most important part of the docket, and a President might be warranted in giving this factor relatively low weight. Unreflective use of statistics obscures this question. More generally, this analysis suggests the need for caution in using invocation figures as a measure of quality, because the figures may be skewed by extraneous factors.

As this discussion of invocation statistics shows, it is important to look beyond the raw data in assessing possible measures of judicial quality. While this is most obvious with the outlier cases in the invocation statistics, it is also true of Choi and Gulati’s best indicator of opinion quality—outside circuit citations.42

C. Citations

Citations are a more promising gauge of opinion quality. There are, however, several possible measures of citation rates to consider. We begin with Choi and Gulati’s “top twenty” citation measure.43 In order to control for the effect of productivity on citation counts, Choi and Gulati identified each judge’s top twenty opinions (measured by citations) and then totaled the cites to those opinions. This measure focuses on the impact of the judge’s strongest opinions rather than looking at all opinions.

To get a sense of the distributions, it is helpful once again to start with a table.

<table>
<thead>
<tr>
<th>OUTSIDE CITATIONS TO JUDGE’S TOP TWENTY OPINIONS</th>
<th>NUMBER OF JUDGES</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-100</td>
<td>2</td>
</tr>
<tr>
<td>101-200</td>
<td>24</td>
</tr>
<tr>
<td>201-300</td>
<td>39</td>
</tr>
<tr>
<td>301-400</td>
<td>19</td>
</tr>
<tr>
<td>401-500</td>
<td>9</td>
</tr>
<tr>
<td>501-600</td>
<td>2</td>
</tr>
<tr>
<td>601-700</td>
<td>2</td>
</tr>
<tr>
<td>701-800</td>
<td>1</td>
</tr>
</tbody>
</table>

42. Choi & Gulati, Empirical Ranking, supra note 2, at 49-50.
43. See id. at 54, 100-03 tbl.E.
44. Figures collated from id. at 100-03 tbl.E.
There are a few things to notice about Table 2. First, unlike the frequency of invocations in Table 1, there are no big gaps in the distribution. While some entries are quite a bit higher than the mean, the numbers climb smoothly to the peak. Second, the distribution is a bit skewed. Third, it is roughly bell-shaped. These impressions are supported by the formal statistics, which show some skew but a kurtosis of close to three (which would characterize a normal distribution).

Why is this a (roughly) normal curve rather than a power-law distribution? Power laws indicate nonlinearity. For example, we would expect a power law if a judge’s top twenty cases were in a feedback loop, where having some of them cited frequently leads to more cites for those cases or for the judge’s other top twenty cases. The absence of such a feedback loop (or at least, of any evidence of such a loop) suggests that among the most prominent cases, citations are not reinforcing but rather are based purely on the precedential value of each opinion.

We get a different picture in Table 3, where we look at total citations, rather than just citations to a judge’s twenty most-cited opinions:

<table>
<thead>
<tr>
<th>TOTAL OUTSIDE CIRCUIT CITATIONS</th>
<th>NUMBER OF JUDGES</th>
</tr>
</thead>
<tbody>
<tr>
<td>101-300</td>
<td>34</td>
</tr>
<tr>
<td>301-500</td>
<td>34</td>
</tr>
<tr>
<td>501-700</td>
<td>22</td>
</tr>
<tr>
<td>701-900</td>
<td>4</td>
</tr>
<tr>
<td>901-1100</td>
<td>2</td>
</tr>
<tr>
<td>1101-1300</td>
<td>0</td>
</tr>
<tr>
<td>1301-1500</td>
<td>2</td>
</tr>
</tbody>
</table>

This does not look remotely like a bell-shaped distribution. Instead, it starts with a large number of judges who get only a few citations and tapers off, at first quickly and then more slowly. This is confirmed by the kurtosis, which is high (5.0) but not nearly so high as the kurtosis for the invocation statistics (which was about ten times

45. If it were completely symmetrical, some judges would have to have a negative number of citations to balance some of the above-average judges.
47. Figures are collated from id. at 94-98 tbl.D.
48. Id. at 99.
This looks a great deal more like a power curve than a normal curve. This should not be surprising—power curves are typical of citation counts in a variety of different fields.

The most notable feature of the power curve is that it tapers off more slowly than a normal curve, so some people get much higher scores than a normal curve would produce. If we assume that the distributions of ability and effort are closer to the normal curve, some process must be boosting the top scores, magnifying differences of ability and effort into large differences in outcomes.

Comparing the last two tables suggests that something of the kind is operating here. The quality of top twenty opinions, which presumably reflects the ability and effort of the judges fairly accurately, follows a normal distribution. But the overall number of citations does not, which indicates that for top judges, differences in citations for non-top twenty opinions are disproportionate to differences in ability and effort. In other words, there is a feedback loop for citations of less significant cases but not for citations of the most significant ones. We could conjecture that significant opinions are cited purely on the basis of their quality, but that less significant opinions are cited partly because the judge who wrote them is well known.

We might also consider the figures for law review citations to judges in Table 4.

<table>
<thead>
<tr>
<th>Number of Law Review Citations</th>
<th>Number of Judges</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-199</td>
<td>14</td>
</tr>
<tr>
<td>200-399</td>
<td>44</td>
</tr>
<tr>
<td>400-599</td>
<td>29</td>
</tr>
<tr>
<td>600-799</td>
<td>10</td>
</tr>
<tr>
<td>800-999</td>
<td>0</td>
</tr>
<tr>
<td>1000-1199</td>
<td>1</td>
</tr>
</tbody>
</table>

49. _Id._ at 107.

50. This feature applies to a broader set of distributions than power curves. Curves having leptokurtosis have steeper peaks and fatter tails than the normal curve, but they have small shoulders—in other words, most cases are either near the average or spread out, with fewer intermediate cases than one would expect. Such distributions are said to be characteristic of situations involving actors with bounded rationality. See Bryan D. Jones, _Politics and the Architecture of Choice: Bounded Rationality and Governance_ 164-84 (2001).

51. Figures collated from Choi & Gulati, _Empirical Ranking, supra note_ 2, _at_ 94-99 tbl.D.
With the possible exception of the top judge (once again, Posner) who seems to be something of an outlier, this distribution does not show strong signs of a superstar effect. Unlike invocations of judges in other opinions, citations to judges in law reviews seem to be distributed among a broader group of judges, with little evidence of the superstar phenomenon. This is all the more interesting because opinions in casebooks show strong signs of the superstar phenomenon, so law professors appear to use different criteria in deciding which cases to use in these two different settings.

Of the various citations figures, which is most relevant for considering potential Supreme Court appointments? The answer is probably the outside circuit citations to the judge’s top twenty opinions. The most important thing a Justice does is write majority opinions. Justices write fewer than twenty such opinions per year—more like ten these days. The tournament covers three years, so the “top twenty” count measures approximately the best seven opinions a judge can write in a year, which is pretty close to the Supreme Court workload. Citation figures for less significant cases tell less about Supreme Court performance. When you get down to the thirtieth least significant opinion a judge can write in a given year, some judges may be much better than others, but this is irrelevant since Supreme Court Justices never write that many majority opinions. In other words, what we care about is the quality of the work the judge does when he or she is really focused on a case and trying hard, not the quality of his or her work on less significant cases.

Another reason to favor the “top twenty” index is that it is more or less a normal distribution, displaying few signs of the superstar phenomenon or other effects that might amplify or distort differences between judges. Invocations are the worst index by this standard, but total citations also show more signs of nonlinear effects than top twenty citations. Law review citations also have some appeal in terms of the absence of apparent nonlinear distortions, provided we are willing to view citation decisions by academics as unbiased and accurate.

In the end, however, the choice of citation basis probably is not critical if we are only looking for a rough measure of opinion quality.

52. See Gulati & Sanchez, supra note 30.
53. Choi and Gulati also collect statistics on citations by the Supreme Court. The problem is that the numbers are small, ranging from 0 to 16, with almost all judges in a small range between 4 and 8. Thus, the numbers could represent mostly random variation. The only judges whose citation counts are listed as statistically significant (at the 95% level) are Posner, Easterbrook, and a group of judges who received zero citations. It would be interesting to examine the Posner/Easterbrook citations to get a better sense of when they are cited. See Choi & Gulati, Empirical Ranking, supra note 2, at 94-99 tbl.D.
54. The judge who tops the list based on this standard is Sandra Lynch. Id. at 100 tbl.E. Lynch replaced Stephen Breyer on the First Circuit.
As Choi and Gulati point out, the correlation between citation indices is very high after correcting for nonlinearities in some of the indices. But if we are looking to pick a winner in a tournament, the choice of index may be more important.

IV. THE RELEVANCE OF OBJECTIVE MEASURES OF PERFORMANCE

Given what we can measure about past judicial performance, are these measurements relevant in evaluating possible Supreme Court nominees? Of the available measures, citation figures seem to be the best indication of professional performance for judges. But of course, they are only an imperfect measure. Indeed, among top judges, differences between citation rates might not even be statistically significant, so picking judges purely on the basis of numerical scores might not be sensible.

There is another reason why the President might give only limited weight to citation figures. I suggested earlier that the most relevant citation figures are probably the number of citations to a judge’s twenty most influential opinions. The reason is that this is comparable to the number of majority opinions for a Supreme Court Justice in any given year. But the most heavily cited opinions by a given judge may not be a random sample of different legal issues.

Based on a recent study of citations to Supreme Court opinions, it appears that the judicial opinions most heavily cited by other judges may tend to involve procedural and other technical issues. For example, in the 1984 Term, the five opinions most cited by other courts involved procedural issues (three cases), ERISA, and an erroneous jury instruction on municipal liability. In contrast, law reviews tend to favor citations to cases involving important social issues. Consequently, only a weak correlation exists between law review citations of Supreme Court cases and citations in law.

This might suggest that the President ought to focus instead on law review citations as better measuring performance in cases involving major social issues. The problem is that these citations arguably may be less reliable as an indicator of quality. Academics do not have the same incentives that judges have to focus their citations on the strongest opinions. They may equally well cite all of the opinions that bear on a particular topic, such as the legality of a specific form of abortion, or they may choose to write about the opinions with

55. See id. at 74. They correct for nonlinearities such as the superstar effect by using the logarithm of citation counts rather than the raw counts. Id. at 72.
56. See Farber, supra note 27.
57. See id. at 870.
58. See id.
59. See id. at 871, 873.
which they most disagree. Moreover, the President might reasonably fear that the political views of law professors would influence their citation practices, which would make law review citations particularly suspect if the President does not happen to share that political orientation. So the President might well find the “top twenty” citation rate to be a more reliable figure.

What this means is that most reliable citation figures probably best measure how well a judge does with technical legal problems. This is surely not an irrelevant consideration for a President. Presumably, it is in the country’s interest to have technical legal issues decided correctly and with the most useful possible guidance to the lower courts. But a President might reasonably be more concerned with how well a judge will handle cases involving major social issues such as abortion, affirmative action, or antiterrorism issues. Citation counts may not give a good indication of the quality of those opinions.

The President might take either of two views about how cases involving major social issues relate to more technical legal cases. On the one hand, he might believe that cases involving major social issues require only the same set of problem-solving and analytical skills involved in more technical cases. Even so, he might still have doubts about relying too heavily on the technical cases as gauges of performance in “hot” cases. He may be concerned about whether a judge who has very strong skills (as shown by citation levels) is able to exclude ideological factors in more political cases. Thus, he would want to look at those cases specifically to see if the judge is able to solve really contentious legal issues through a dispassionate application of legal skills.

On the other hand, the President may believe that decisions in “hot” cases involve something more than the application of technical legal skills—that personal values or empathy or statesmanship are especially important in these cases. Citation rates in cases with more legalistic issues will only weakly reflect the presence of these values. Whichever view the President takes about the relationship between technical cases and “hot” cases, measuring the quality of opinions in the more technical cases would give him only limited guidance.

So far, I have been assuming that the President finds it irrelevant how the judge will vote on particular legal issues but is only looking for general indicators that a Justice will have the character and skill to make good decisions in hard cases. But it is not clear that a con-

---

60. This may be part of the explanation for why individuals with federal circuit experience have not historically tended to rank as “great” Justices, as observed by Steven Goldberg, Federal Judges and the Heisman Trophy, 32 FLA. ST. U. L. REV. 1237 (2005).
scientious President should take that view.\textsuperscript{61} There are two reasons why the President might give weight to his view of how the Justice will vote on key issues.

First, if the President has sufficient faith in his own views of the constitutional merits of these issues, he might view a judge's ability to reach the "right answers" to be a critical test of judicial ability and character. A judge who keeps getting the wrong answers must be doing something wrong, even if the President does not feel confident of his ability to determine the exact nature of the deficiency.

Second, the President might view the outcomes in those cases as having a social significance that outweighs any question of technical legal correctness. He may believe that abortion is a form of mass murder or, alternatively, that access to abortion is critical in order for women to be equal citizens. It might be wrong for a judge to base a decision entirely on her views about such social and moral considerations. But the President is not a judge and would seem to have the discretion to nominate the judge whom he thinks will be, all things considered, best for the nation's welfare.

The point is simple: If the President simply wants to make the best possible appointment, general measures of professional competence are not likely to be the exclusive factor in the decision. There might, however, be some collateral reasons for excluding or at least minimizing other factors.

To begin with, the President might believe that introducing other factors undermines confidence in the courts by making them appear politicized. This is certainly a concern, and one that has been frequently voiced. But nomination controversies do not seem to have harmed the courts' reputation to date, perhaps because few people pay attention to the appointments process.

Alternatively, the President might believe that basing his selection on apparently objective measures will lead to a smoother appointments process, with less unseemly wrangling with Congress. This would surely be desirable. But it is another question whether it is realistic, given the country's current political polarization. And in any event, the tournament might not be successful in reducing partisanship. For instance, the putative winner of the tournament, Judge Posner, has taken controversial positions which would be offensive to important political factions. It seems unlikely that their opposition would be silenced by the strength of his professional credentials.

Or perhaps the President might believe, like Choi and Gulati, that using objective selection criteria would give lower court judges an in-

\textsuperscript{61} By conscientious, I mean a President who wants only to make the best appointment for the good of the country, excluding all partisan or personal considerations.
centive for high performance. Here, there are two questions. First, how strong would the incentive effect be? Only a few judges would be serious contenders for victory in the tournament, and the incentive effect would be strongest for them—but presumably the best performers are least in need of additional incentives. Second, is the gain in incentives for lower court judges worth making an otherwise less desirable Supreme Court appointment? Looking at the welfare of the country as a whole, the President may think that small changes in the performance of lower court judges are not valuable enough to be worth the sacrifice.

V. CONCLUSION

Choi and Gulati propose a tournament, whereby promotion to the Supreme Court would be based on objective measures of a circuit judge’s performance. Although it is intriguing, such a tournament is not, at least at present, a plausible option. Of the three traits that Choi and Gulati seek to measure, one (productivity) is ambiguous as an indicator of quality, and another (independence) is hard to measure and ideologically charged. Citation counts—and in particular, Choi and Gulati’s “top twenty” counts—provide a better metric. This measure of opinion quality is imperfect but seems relevant to the President’s selection.

Yet, even if overall judicial performance could be perfectly measured, the best lower court judge might not be the best choice for the President. Presidents are legitimately concerned with other factors, including how the nominee might vote on current key issues and how the nominee is likely to approach future national controversies. The tournament idea puts too much weight on one dimension, albeit an important one.

Although the tournament may not be “ready for prime time,” collecting and disseminating data on circuit judge performance could be very useful. For example, a website could post the best available measures of judicial quality.62 If it established a reputation for objectivity, such a website could be a valuable resource during Supreme Court vacancies. It would be up to the relevant actors—Presidents, Senators, press members, and the public—to decide on the relevance of the information. More complete information could well improve the quality of the debate.

Furthermore, such a website also could serve less dramatic, but still significant, secondary functions. It would allow law students to

---

make better-informed decisions about clerkships. Even without the prize of a Supreme Court nomination, simply posting the measures could also create some useful incentive effects. Finally, the website would be a valuable resource for scholars.

It would be terrific to reform the acrimonious Supreme Court appointment process, but this may be too much for mere law professors to achieve. If Choi and Gulati manage to have only a small incremental effect on judicial selection, that will be noteworthy. In the meantime, the data they have already collected will provide scholars with significant insights into patterns of influence among appellate judges.
I. INTRODUCTION

The lure of doing empirical research in legal scholarship is strong, but the appeal of critiquing judicial selection is stronger. So it is rather surprising that empirical analysis figures very little in the voluminous legal commentary on judicial selection. Instead, legal scholars devote little attention to the actual numbers in judicial selection; they generally defer to, or merely accept, the empirical work.

---

* Arthur B. Hanson Professor of Law, William & Mary Law School; Visiting Fellow, James Madison Program in American Ideals and Institutions, Princeton University, Spring 2004. I am grateful to Scott Baker and Mitu Gulati for helpful comments on an earlier draft. My contribution to this Symposium is an expanded version of my essay entitled Merit vs. Ideology, 26 CARDOZO L. REV. 353 (2005).
done by others, particularly social scientists, on judicial selection. While some of these studies are excellent, many have problems. In this Essay, I explore three basic challenges to doing empirical studies of judicial selection and suggest ways to meet these challenges. My hope is to give some useful guidance to legal scholars and others interested in developing a more sophisticated understanding of the judicial selection process through empirical analysis.

The first challenge, discussed in Part II, is defining merit. Many legal scholars rarely discuss merit in their commentaries on judicial selection, even though developing a coherent notion of merit is essential for evaluating the quality of judicial nominations. Consequently, in Part II, I examine the empirical challenges posed by several plausible definitions of merit. This discussion should help to illuminate, inter alia, the importance of merit as a normative criterion for judicial selection as well as the significant relationship between merit and the one thing almost all commentators assume is a principal driving force in judicial selection: Ideology, by which I mean pre-commitments to certain outcomes or approaches for analyzing constitutional issues, regardless of the facts of particular cases.

In Part III, I examine the benefits of avoiding the popular practice of labeling judicial nominees on the bases of their supposed ideologies. Politicians and social scientists share a strong proclivity to pigeonhole judges and Justices based on their supposed ideologies, but their labels tend to be overly simplistic, inaccurate, incomplete, misleading, and value-laden. They tend to obscure, if not ignore, the subtle differences in judicial philosophies as well as the basic fact that some Justices and judges do not have fixed ideologies either at the time of their initial appointments or perhaps ever. Moreover, because of ideological drift (and other factors), the categories scholars and others use in analyzing judicial performance are not static. Perhaps the most serious problem with labeling is that it obscures the middle, or the mainstream, in American constitutional jurisprudence. Politicians fight to place their nominations within the mainstream for a reason: It provides a benchmark against which to measure politicians' and others' characterizations of judicial performance. Each side strives to include its nominees within the mainstream and to push the other side’s outside of it. Moreover, each side has a strong incentive to push the envelope in confirmation proceedings, because it will help expand the safe ground available to its nominees but

---

situate the other side’s nominees closer to, if not beyond, the outside of the mainstream.

In Part IV, I explore possible empirical tests for determining the relative impacts of almost a dozen factors on the fates of judicial nominations. These recommendations will help legal scholars (and others) move beyond their unfounded assumptions about the forces driving federal judicial selection and be more precise with the terms they employ for characterizing judicial qualifications or performance. While it may never be possible to bridge the divide among national political leaders on such basic things as defining merit or ideology in judicial selection, empirical research at least holds the promise of illuminating common ground for those outside the process to evaluate what happens inside of it.

II. MERIT

In this Part, I examine a basic question that presumably is of great interest to everyone concerned with the quality of judging: How do we measure fitness for office and, particularly, how do we determine who are the best qualified people for appointment to the U.S. Supreme Court, the U.S. courts of appeals, and U.S. district courts? There are several possible definitions of merit, each of which poses a challenge for empirical analysis of judicial selection. I discuss several possible definitions of merit and the respective empirical challenges posed by each of them.

A. The Conventional, Elitist Definition of Merit

One way to define merit is simply to examine how it has been defined in the past. A review of the critical literature on merit indicates that, as a descriptive matter, critical elites have always defined merit to suit their own purposes. Different sectors and professions develop their own, self-serving conceptions of merit. With respect to judicial appointments, there has never been some objective, or neutral, criterion of merit. Instead, the governing elite has made judicial appointments to further its own interests. Consequently, merit is defined so as to allow for, if not to maximize, the appointments of relatives, friends, and especially political allies. For any given presidency or era, the essential tasks are to figure out the conception of merit advanced by the governing elite and how well nominees fit this


particular conception. It has been quite common for Presidents to choose judicial nominees who not only share their constitutional vision but are also likely to be supportive of their administrations’ policies. For instance, President George Washington chose his judicial nominations based in part on their support for the ratification of the Constitution and commitment to Federalist ideals, while President Lincoln opted for judicial nominees who supported the Union and Reconstruction.

There are, however, three major problems with relying on an instrumental conception of merit as an empirical measure of the quality of judicial nominations. First, the conception is unstable. It changes at least as often as the governing elite does. For instance, President Reagan, from 1980 through 1986, employed a conception of merit that included rigid commitment to original understanding. Once Democrats took control of the Senate in 1986, the pressure mounted on President Reagan to moderate his definition. He soon did: After his first two nominations to replace then-retiring Justice Powell failed, President Reagan modified his conception of merit to settle on someone with more moderate views, Anthony Kennedy. President George H.W. Bush might have been disposed to employ the same definition of merit adopted originally by President Reagan. But facing a Democrat-controlled Senate, with a majority hostile to rigid commitments to original understanding, President Bush chose a candidate without a paper trail on ideology—David Souter—to replace then-retiring Justice Brennan.

Second, it is a mistake to assume the governing elite is candid about its conception of merit. In some cases, governing elites are not forthcoming—and some might go so far as to suggest they are deceptive—about their conception of merit. President George W. Bush insists that merit is his principal selection criterion, but he has never given a detailed definition of his conception of merit. Nor does merit, however defined, seem to explain all his judicial nominations. For instance, President Bush twice nominated and eventually made a recess appointment of Charles Pickering, Sr., to a seat on the U.S. Court of Appeals for the Fifth Circuit, though Pickering was the most reversed district judge in his circuit. It is hard to explain Pickering’s appointment as based on some objective criterion of merit.

Third, the governing elite’s definition of merit works very poorly as a normative criterion. Using each administration’s definition of merit to evaluate its nominees allows it, not any external or inde-

5. See id. at 86-92.
6. See, e.g., Johnsen, supra note 2.
pendent source, to define the terms on which they may be measured. Each administration’s definition of merit is likely to be self-serving, so in all likelihood it would be a rare case—perhaps like Pickering’s nomination—that does not fit within its definition of merit. The less precise an administration is in defining merit, the easier it will be to show that its nominees satisfy it. Thus, it is conceivable that even Pickering can be defended as a merit appointment on some grounds, such as his longstanding experience as a district court judge and the respect that he has earned from the lawyers appearing before him as well as the members of local and state bar associations. Yet, the less precise an elite’s definition of merit, the less useful it becomes as a normative criterion, because it is not stringent enough to rule anything out. Moreover, it would be extremely unfair to use one governing elite’s definition to measure the quality of an appointment made by a different governing elite, because they will have been likely defined in opposition to each other. President Reagan’s definition of merit, for instance, was deliberately designed to contrast with the appointments of his Democratic predecessors as well as the politically moderate Gerald Ford. He chose his definition in part because it ruled out the likelihood of choosing people like those nominated to judgeships by Ford or Presidents Johnson and Carter.

Thus, the elite’s definitions of merit are of limited relevance to empirical research. To be sure, they may be useful for clarifying a particular administration’s expectations of its nominees. But elitist definitions of merit are unlikely to be useful in satisfying more ambitious empirical objectives. Empirical research requires developing criteria with more explanatory potential and normative value.

B. The Public as Arbiter of Merit: Social Scientists Do Not Have a Uniform Conception of Merit

Some agree that merit is a purely instrumental concept. Perhaps the most aggressive advocates of this conception are Harold Spaeth and Jeffrey Segal. They argue that judicial selection is not based on any neutral conception of merit. Instead, they claim, Supreme Court Justices reflect the values, or preferences, of the governing elite. They are chosen, in other words, because they are likely to uphold the values and preferences of the people who appoint them. Understanding judicial selection in this manner helps explain why Supreme Court Justices never stray far from standing behind the policies or cultural beliefs of the people who appointed them. Instead, Segal and Spaeth argue, Justices are picked on the basis of their

[7. See Jeffrey A. Segal & Harold J. Spaeth, The Supreme Court and the Attitudinal Model Revisited 86-97 (2002) (describing judges as motivated in their decisionmaking primarily by personal policy preferences).]
likely ideologies with respect to social policy. Once on the bench, they merely vote their policy preferences.

Many other social scientists take contrary viewpoints on merit. They believe that in empirical research it is important to ground a conception of merit in a source independent from, or external to, the actual judicial selection process. An external definition of merit is useful normatively, because it provides an arguably neutral benchmark detached from an actual Justice’s selection against which to measure his or her qualifications and performance. For instance, social scientists sometimes use the public as an external source for defining merit. This definition allows researchers to measure judicial selection against public opinion.

In this Symposium, Lee Epstein and her coauthors suggest a variation on this approach. They suggest the media as one possible evaluator of merit. They consider merit as simply being whatever the media has defined it to be. This is, however, a problematic measure for several reasons. First, it is unclear why the media should be regarded as neutral on this matter. Indeed, many thoughtful people believe that the media is not neutral. They believe that the media has its own ideological agenda, which is likely to influence its coverage of events and its evaluations of the merits of judicial nominees. For instance, it is hard to imagine that The Wall Street Journal editorial page or the Fox News Network will seriously question President Bush’s declarations about the merits of his judicial nominees.

Second, the media has no apparent interest in being or becoming a neutral arbiter of merit. There is no good reason to believe that the media has the public interest at heart and thus can be fairly or sensibly viewed as acting on its behalf. By all accounts, the media loves to cover scandal and conflict. The need for the media to make money, coupled with intense competition arising from the twenty-four-hour news cycle and the Internet, has exacerbated the declining interest in simply reporting the hard news—that is, facts and figures. The media increasingly prefers to report soft news—that is, speculation and commentary. Drama is likely to attract reader or viewer interest, and so the media increasingly reports as much drama as it can. A neutral assessment of merit, if it were to be made at all, is likely to be lost in the avalanche of soft news that dominates media reporting. Moreover, many commentators are not lawyers, and many of the

---

ones who are lawyers are unlikely to be neutral—they are likely to be fierce combatants for one extreme viewpoint on a nomination.

Third, the media lacks any expertise in evaluating merit. It is unclear why it should be supposed that the media generally has any special insight into the qualifications for judicial service. Some commentators have law degrees and may have practiced law or served in government; however, most are not trained as lawyers. Moreover, many professions evaluate merit based on the assessments of those within the profession. Some quickly write this off as merely a custom or price of admission into a particular guild. While it may be true that professionals are not necessarily the only ones who can evaluate what they do professionally, it does not follow that professionals are ill equipped to evaluate each other’s professional performance. Thus, yet another important empirical study to undertake is measuring the extent to which a particular profession, such as legal practitioners or judges, is well equipped to evaluate its own practitioners.11

C. Assumptions on Which Most People Agree

One possible reason legal scholars do not discuss merit is because most of us—indeed, perhaps most people—share a number of assumptions about merit in judicial selection. It is possible that with respect to merit we might have more consensus than generally acknowledged. Thus, it is important to consider the assumptions about merit on which there may be substantial if not uniform agreement. Below, I consider two different ways for illuminating such assumptions and the problems with empirically demonstrating each, particularly separating merit from ideology.

1. Imagining the Ideal Judicial Nominee

Imagine, for a moment, that you have been asked by the President to draft a list of qualifications for nominees to the Supreme Court and to other Article III courts. Imagine further that you do not know which particular President has made this request. You are, in other words, behind a Rawlsian veil of ignorance,12 for you do not know anything particular about whether this will ultimately help or hurt you, what kinds of cases will likely come before the courts, the President’s party, and the composition of the Senate. Is it possible to draft such a list, and if so, what qualities would be on the list? Below, I consider five possible qualities that could conceivably satisfy our collective assumptions about merit.

(a) Professional Experience and Accomplishments

It is easy to see the importance of meaningful, first-hand professional experience to judicial selection. We want our judges and Justices to be familiar with the business and responsibilities of the courts to which they have been nominated. We do not want appointees to need on-the-job training. The ideal is someone whose experience specially qualifies her or him for judicial appointment.

The requisite experience might not be the same for every Article III court. Presumably, the ideal experience for a Supreme Court Justice is practicing before the Court and perhaps serving in other offices, such as lower court judgeships, that employ similar (but not identical) skills. Meaningful experience might also include serving in a significant public office, for it might enrich the nominee’s understanding of the system from which the laws appealed to her Court will come. Rich professional experience is bound to sharpen nominees’ judgments and provide a solid foundation from which to approach the significant legal questions that come routinely before the Supreme Court.

The ideal experience for a federal court of appeals judge is substantial appellate practice or prior service as a judge on a comparable court or perhaps on a federal trial court. Meaningful experience might also include demonstrated or proven expertise in some areas of the law. The ideal may be that the leading practitioners or scholars in different fields would bring their respective expertise (and the skills that allowed them to develop it) to the task of judging.

The ideal experience for a district judge is substantial trial experience. Not all trial lawyers will make good district judges, but all good district judges presumably understand how to manage trials of all shapes and sizes and large caseloads. The ideal experience may also include service as a judge on a comparable court, such as a state supreme court.

Demonstrating empirically the extent to which judicial nominees have, or deviate from, ideal experience should not be difficult. We already have considerable data on the professional backgrounds of the people nominated to judgeships. For instance, we have the rates of affirmance and reversal for sitting judges. We can also count years devoted to legal practice and determine the relative weight to be assigned to the professional work nominees did prior to their nominations. Judicial nominees’ professional experience need not have been in the public sector, but the more firsthand experience with the legal system a nominee has, the better. We could assign more weight to public service experience, but we need not discount entirely other professional experience. Furthermore, we could measure familiarity
with the legal system based on the data gathered by bar groups or those who have worked closely with the nominees.

(b) **Legal Acumen**

We expect judicial nominees to have very high degrees of legal acumen. We expect judicial nominees to be highly intelligent, perhaps to have performed quite well in law school, maybe even to have attended an elite law school. At the very least, we would want to make sure that the nominee has very sound legal skills; is capable of asking intelligent, probing questions; thinks critically (if not imaginatively) about legal problems; identifies legal issues in a wide range of problems; is trained at problem-solving; is diligent; and understands the special duties that she will be called upon to discharge.

Measuring this criterion entails coordinating different data. There will be some objective data (such as the law school attended or the number of law review articles written) and some subjective data (such as measuring the opinions of the bar or those who have practiced before or with the nominees).

(c) **Judicial Temperament**

We expect judicial nominees to have excellent judicial temperament. The ideal temperament for a judge or Justice is presumably to have the capacity to make decisions evenhandedly, to be open-minded in listening to and considering the arguments in the cases that come before him, and to be respectful to litigants and other appellate judges or Justices with differing opinions. A district judge may largely work alone in deciding cases but still needs a great deal of patience to sit through long trials and other legal proceedings. Moreover, a good judicial temperament requires, of course, a disposition to follow the law. Judicial nominees need, in other words, to demonstrate in some ways that they are well suited to resolving legal disputes rather than rewriting the laws they are interpreting. Judicial nominees also need to be able to handle the intense pressures that come with the responsibilities of being judges or Justices.

Collegiality is a related measure of ideal judicial qualifications.\(^\text{13}\) Collegiality requires getting along with the other Justices or judges with whom one must work. Collegiality also entails being able to build coalitions and to maintain cordial relations with other judges or Justices, regardless of the extent to which one may agree or disagree with their views in particular cases. Maintaining cordial relations is

no mean feat on the Supreme Court, which Justice Holmes once described as “nine scorpions in a bottle.”¹⁴ Not all people who must work together in relatively close quarters successfully maintain respect and civility over long periods of time, but the ideal nominee must have this capability.

Empirically demonstrating excellent judicial temperament or collegiality is not easy. To be sure, if nominees have had lots of verified complaints filed against them for temperamental outbursts, those would be signs of problems with temperament. If the complaints come from other judges, they undoubtedly will call into question a nominee’s collegiality. Moreover, taking temperament or collegiality into account invites unsubstantiated rumors or insinuations from critics. Measuring nominees’ collegiality or temperament thus requires sorting fact from fiction. It requires, in other words, developing credibly neutral measurement.

(d) Writing Skills

Yet another criterion for the ideal nominee is excellent writing ability. This is particularly true for Supreme Court Justices and federal appellate judges, who are called upon to draft large numbers of opinions. The ideal Supreme Court or federal court of appeals nominee should be able to write clear, coherent opinions. It is especially important that the ideal nominee have the ability to craft opinions reflecting multiple viewpoints. Moreover, it is important for the nominee to be able to compose opinions relatively quickly given the time pressures under which judges and Justices operate.

Measuring writing ability involves both objective and subjective data. The objective data include judges’ productivity and the amount of time it takes for judges to produce opinions, particularly as compared to the other judges with whom they regularly sit. The extent of a nominee’s nonjudicial writings may also be pertinent. It reflects, inter alia, their contributions (if not their leadership) in refining our understandings and reforms of different areas of the law. Subjective data include the quality of a nominee’s writings and speeches. The quality of the latter depends on the assessments of others, including, but not limited to, peers and experts.

(e) Integrity and Character

Integrity is essential to the ideal judicial nominee. A judicial nominee’s integrity must be beyond question in order for her or him

to exercise the fragile moral authority of a Supreme Court Justice, federal court of appeals judge, or district judge. Judges and Justices embody the law, and they need to comply with the very laws they expect others to follow. If nominees have not followed the laws they expect others to follow and for whose violations they may sentence other people, they no longer can claim the special moral status on which their job depends.

Closely related to nominees’ integrity is their moral character. Stephen Carter and Larry Solum are just two of the many scholars who insist that a Supreme Court Justice ought to have a strong, moral character. At the very least, having a strong, moral character means having the courage of one’s convictions and the strength not to alter one’s opinions, or decide cases, for the sake of public or peer esteem.

Integrity and character are easier to measure in their absence than they are positively. For instance, criminal convictions would almost certainly disqualify someone from judicial appointment, though it is conceivable that if the crimes were misdemeanors and committed many years before the nominations they might be dismissed as irrelevant. Admissions or proof of wrongdoing further demonstrate problems with integrity or character. The fact that Justice Fortas continued after his appointment to give policy advice to President Johnson raised a question about his professional judgment. Even worse, his accepting money from a convicted mobster was an ethical lapse, just as it is unethical for a judge or Justice to rule on cases involving former clients or companies in which he or she owned stock. Complaints made by spouses or children in custody disputes might raise questions about nominees’ characters; even if no criminal conduct were alleged, charges might be raised about the person’s moral judgment.

In the absence of documentation, integrity and character are established through subjective data. They will be based largely on testimonials from other people. Senators tend, however, to give more weight to negative appraisals. Because most nominees receive positive testimonials, the latter end up not receiving much attention unless they have come from people whose opinions carry great weight with Senators. In contrast, negative testimony draws much greater

16. For an account of these interactions, see Laura Kalman, Abe Fortas 310-18 (1990).
attention. It makes headlines, provokes lots of questions, and raises potentially fatal doubts.

(f) Other Factors

There are other qualifications that ideal nominees arguably need to satisfy. Besides the factors already mentioned, Presidents might also be interested in nominees’ religion, ethnicity, gender, and health. These other factors might be important in diversifying the composition of the federal courts or satisfying underrepresented segments of society. Moreover, nominees’ ages have been very important to some Presidents who wanted to ensure that their appointees could continue to serve as judges and Justices long after they left office.

(g) Empirical Complications

Of course, the criteria that are relevant for determining ideal nominees are one thing, while the things at which Presidents or their advisers might look in order to measure actual nominees are quite another. The values of those charged with selecting a nominee will inevitably influence what they choose to look at and how they will perceive it. Moreover, it might simply be unrealistic—or reckless—to ignore factors such as timing, the President’s party, the composition of the Senate, the nominee’s political or party affiliation, or the compositions of individual federal courts. For instance, the composition of the Senate might be quite pertinent to a Supreme Court nominee’s chances of confirmation. Indeed, a President might be inclined to choose different people, depending on whether his party controls the Senate or whether the minority has enough members to filibuster a contested nomination. Certain factors are bound to complicate the nominating process. For instance, the proximity of the next presidential election cannot be ignored, since the opposing party has successfully rejected or delayed more than a few Supreme Court nominees in the hope of preserving the vacancies for Presidents from their party.18 And we have not yet mentioned a nominee’s likely ideology or how well a potential candidate interviews for the job as possible complicating factors.

Nor have I yet acknowledged other, arguably more serious problems with collective assumptions about merit. First, we still have to demonstrate empirically our collective assumptions about merit. It is one thing to assume we share some consensus on merit, but we need to show precisely what, if anything, our shared assumptions are. This requires figuring out whose opinions ought to count and how to

18. See ABRAHAN, supra note 4.
figure them out. Second, even if there were shared assumptions about merit, there may not be any consensus on what kinds of activities or data satisfy our shared criteria for merit. For instance, it is possible there is relatively widespread consensus on the relevance of professional experience to judicial selection but not on which experiences ought to count or on the relative weight (or relevance) of different kinds of experience (say, public defender versus federal prosecutor). Third, there may not be any consensus on the relative weight of different factors, such as experience and judicial temperament. People might have reasonable differences of opinion about how to prioritize different factors or whether one factor ought to count more than another. Nor may there be any consensus on the significance of an absent factor or how much weight to attach to a negative rating with respect to a particular criterion, such as judicial temperament. Again, people might have reasonable differences of opinion about what needs to be shown in order to disqualify a nominee altogether.

The large number of potential considerations helps to explain why some Presidents, or their advisers, might prefer to break the nominating process into first- and second-order selection criteria. The first might allow for a relatively sizeable list of potential candidates, while the second might be used to cut the list down to size (if not down to one). Interviews might be used to cut a narrowed list even further (at which point, of course, a great deal depends on who is doing the interview, the questions asked, and the nominees’ responses).

It is possible that recognizing the large number of potential considerations discourages academics from pondering the qualifications of ideal nominees to the Court and other Article III courts. Academics might view such an exercise as futile, for they appreciate that judicial nominees are not chosen in vacuums. Yet neither Senators nor academics hesitate to evaluate nominees, particularly those to the Supreme Court, on the basis of some criteria. The question thus remains as to the appropriate criteria for measuring the quality of a particular nomination. Of course, the fact that a nomination falls short of an ideal is not necessarily an argument against it. Supreme Court nominees usually enter the confirmation phase with at least a presumption, or likelihood, that they will be confirmed. It thus usually takes some rather significant things—not just some deviation from an ideal—to put nominations in trouble. Nevertheless, the stronger nominees’ credentials (or the more closely they approximate an ideal), the tougher it may be to undermine their nomination. Thus, looking at another way to determine ideal credentials

might be fruitful for providing at least one significant measure for evaluating the relative strengths of particular nominations.

2. Determining Merit in Reverse

Part II.C.1, supra, examined possible judicial selection criteria on which there might be consensus. This section considers determining qualifications by looking at merit in reverse. It considers whether it is possible to infer from the Justices we might generally agree were “great” or “excellent” what they might have had in common prior to their appointments. The question is whether the signs of at least potential “greatness” or “excellence” were evident at the times of the appointments of Justices who later proved themselves to be first-rate Justices.

I illustrate this tack through two examples. The first is Justice Benjamin Cardozo. His name appears on most lists of great Justices, so the question naturally arises whether, or in what ways, Justice Cardozo’s greatness or excellence was evident at the time of his nomination. Throughout his career—first as a lawyer specializing in appellate briefs (from 1891 to 1913), then as a judge (since 1914) and later chief judge (since 1926) of the New York Court of Appeals—Cardozo, nominally a Democrat, had enjoyed the confidence of all political factions. Achieving this level of confidence was especially significant because he had done it in an era when state courts (and his court especially) were widely revered. Cardozo was also the author of several highly regarded books and had received honorary degrees from many universities, including Yale, Columbia (his alma mater), and Harvard. Many of his decisions in such areas as torts and contracts influenced judges and courts throughout the nation. Thus, he evidently had, by the time of his appointment, compiled ample judicial experience, shown considerable legal acumen, and demonstrated excellent judicial temperament, collegiality, and leadership on a leading court. At the time of his appointment, Cardozo’s integrity and character were beyond reproach.

My second example involves another New Yorker, Charles Evans Hughes, whom many believe was a first-rate jurist (not once but twice!). When Hughes was first nominated and confirmed to the


22. For a brief account of Cardozo’s legal training at Columbia, see id. at 40-50.

Supreme Court in 1910, he already had outstanding credentials. He had been an active practitioner with one of the leading law firms in the country, been a leader of the New York and national bars, and devoted himself to substantial public service. At the time President Taft appointed Hughes as an Associate Justice, Hughes was serving with distinction as the Governor of New York. As an Associate Justice, Hughes authored a number of significant opinions, demonstrated respect for his colleagues and opposing arguments, and displayed an evenhanded temperament. After leaving the Court six years later to run unsuccessfully for President of the United States, he served as president of the American Bar Association; argued several cases successfully before the Supreme Court (and performed significant pro bono work); served for four years as Secretary of State under Presidents Harding and Coolidge (1921-1925); and served on the Permanent Court of International Justice, or World Court (1928-1929). Few nominees to the Court have ever matched Hughes’ record of public service prior to either of his appointments to the Court, and fewer had records of public service that commanded the respect of the leaders of both parties (though this did not save him from having a significant minority of Senators vote against his nomination as Chief Justice for fear of his allegiance to big business). Hughes brought statesmanship to the task of judging.

My point is not to suggest that either Cardozo or Hughes ought to be the ideal model of a Supreme Court Justice or Chief Justice. Rather, my point is that if we are sincerely interested in measuring merit, we might consider inferring from certain nominees’ records, during their respective appointments, appropriate criteria for meritorious appointments to the Court. But we rarely do. This may be because we rarely take merit into consideration without some reference to ideology. Many, if not most, of us might suspect that most Presidents and Senators are preoccupied with ideology in assessing judicial nominees. Consequently, we need to consider the implications of the linkage of merit to ideology in the federal judicial selection process.

25. See id. at 68-71.
27. Hendel, supra note 24, at 74-77.
28. 2 Pusey, supra note 26, at 640-47.
29. Hendel, supra note 24, at 78-90 (discussing the debate over Hughes’s nomination to the position of Chief Justice).
D. Assumptions on Which Few People Agree

A fourth way to define merit is through assumptions on which few people agree. It is possible that many, or even most, of us might agree that ideology is pertinent to judicial selection, though we might not agree on the particular ideology that we would prefer for judicial nominees to have.

Indeed, no one seriously thinks that President George W. Bush has been using the same criteria that President Clinton employed in choosing which people to nominate to district and circuit court judgeships. Instead, we strongly suspect (based on leaks and outcomes) that President Bush is considering different sets of people from those that President Clinton considered nominating as judges. The differences in their nominees go beyond party affiliations or allegiances; they reflect differences in experience, political commitments and service, and attitudes about how to decide constitutional cases. These attitudes are what some people might call ideological commitments.

Yet it is reasonable to wonder whether there are any selection criteria on which Presidents Bush and Clinton (or their respective advisers) would agree. (Presidents Clinton and Bush apparently did agree on two nominees—Judges Roger Gregory and Barrington Parker, Jr.—whom they nominated and who were ultimately confirmed by the Republican-led Senate in 2001.30) Presidents Clinton and Bush (and their supporters and advisers) each have claimed that they nominated the best-qualified people as federal judges, but these claims beg the question: how do we determine merit, or who are the best-qualified people for judicial appointments? It is not immediately clear why or how both Presidents could be appointing the best-qualified people given that they appear to have been nominating quite different kinds of people to judgeships—people with different backgrounds, political experience, party affiliations, sponsors, and attitudes.

Given these circumstances, a typical refrain from scholars is to insist that ideology matters, because it frequently makes the critical difference in whom the President nominates or whom the Senate confirms to Article III courts, and thus we need to focus on the likely ideologies of judicial nominees in evaluating whether they ought to

---

30. In his final year in office, President Clinton nominated Roger Gregory and Barrington Parker, Jr., to the U.S. Court of Appeals for the Fourth Circuit and for the Second Circuit, respectively. The Senate never acted on those nominations. Consequently, President Clinton in his last month in office designated both nominees as recess appointees to their respective courts of appeals. These appointments would have expired at the end of the next congressional session. However, in March of his first year in office, President George W. Bush announced his first set of nominees to the federal courts of appeals, including Gregory to the Fourth Circuit and Parker to the Second Circuit. David G. Savage, Bush Picks 11 for Federal Bench, L.A. TIMES, May 10, 2001, at A1.
be confirmed. Walter Dellinger’s proposed solution to the impasse over some of President Bush’s judicial nominees has the distinct virtue of smoking out whether ideology is what matters most to each side.31 He proposes that each President ought to agree to nominate at least one of a preselected few people approved by the opposition party in exchange for a relatively smooth confirmation process for every three or four people he prefers to appoint to a particular circuit court of appeals. For instance, in exchange for his getting Miguel Estrada, Bret Kavanaugh, and Tom Griffith appointed to the U.S. Court of Appeals for the District of Columbia,32 President Bush would be obliged to nominate someone from a group of potential nominees approved by the Democratic caucus in the Senate. If the President were to refuse, Dellinger argues, then it can only be because he clearly prefers to zealously protect his prerogative to take ideology into account in nominating judges.

Dellinger poses a powerful test of presidential commitment to ideological criteria for judicial nominations. Nevertheless, it is possible that Presidents may zealously protect their nominating authority (as did, for instance, Presidents Tyler and Madison in the nineteenth century33) for reasons other than the desire to ensure the ideological purity of their judicial nominations. Presidents may wish to preserve their autonomy to nominate people to judgeships for such other reasons as rewarding personal or party fealty, currying the favor of particular Senators or constituencies, and broadening the diversity of the federal judiciary. Of course, none of these reasons for appointment is mutually exclusive from fulfilling certain ideological criteria. It is possible that Presidents, or at least their advisers, might define merit as an additional criterion for nomination or perhaps as the critical factor for choosing among potential nominees or for determining the potential sets of nominees for particular judgeships. Indeed, some Presidents, or their advisers, might define merit as including a particular ideological orientation with respect to constitutional interpretation. It is not unprecedented by any means for Presidents to select people as nominees based on the extent to which the nominees conform to the presidents’ notions as to the duties they expect judges or Justices to perform. President Reagan, I suggested earlier, seems

32. Miguel Estrada, Bret Kavanaugh, and Thomas Griffith are three Bush nominees to the U.S. Court of Appeals for the District of Columbia on which the Senate never acted. Democrats successfully blocked a floor vote on Estrada’s nomination, while the Judiciary Committee, as of the date of our Symposium, never acted on either Kavanaugh’s or Griffith’s nomination. I do not know Estrada personally, but I do know both Kavanaugh and Griffith, both of whom very kindly and generously have given their time to visit, more than once, my constitutional law classes.
to have defined merit, at least in part, as including certain ideological commitments.\textsuperscript{34} He insisted that his staff and Senators recommend candidates for judicial nominations that fit particular criteria, including a rigid commitment to original understanding in all cases. Thus there are administrations that define merit in terms of ideology.

The fact that administrations—and a number of commentators—define merit in terms of, or in part based on, ideology leads some empirical researchers to do the same. For instance, Segal and Spaeth measure the attitudes, or ideologies, of judges and Justices based on newspaper editorials at the time of their respective appointments,\textsuperscript{35} while Lee Epstein and Gary King have suggested determining nominees’ ideologies based on the Senators sponsoring them.\textsuperscript{36} But Senators sponsor nominees for many different reasons, not the least of which is payback for political support or fealty. Some Senators also may not care about ideology or make assumptions about nominees’ ideologies based on what others tell them. Moreover, of course, as others have suggested, one of the nation’s most important appellate courts—the U.S. Court of Appeals for the District of Columbia—is not within a state. The District of Columbia has no Senators, and thus senatorial courtesy does not work in the same way for its judges as it does for those from the fifty states. An additional complication is that some states do not have any Senators from the President’s party, in which case they have no input on nominations or the President pays more attention to what the highest elected official within the state from his party has to say about prospective nominations. In short, it is possible that sponsoring Senators signal nominees’ possible commitments to particular ideologies, if and only if the Senators are well known (and can be shown) to prefer nominees with such commitments.

Empiricists should also consider three other possible measures of ideology. The first is simply to look at how different organizations characterize nominees’ judicial attitudes. These definitions are unlikely to be neutral; they are as likely to be self-serving as the assessments of any other parties with vested interests in the fates of particular judicial nominations. A second measure might be reflected in the nominees’ activities with organizations whose members are known to have or share particular attitudes about constitutional law. The problem with this standard is that it is unclear how many or even which activities demonstrate ideological commitments that will carry over into judging. Moreover, some nominees might not have

\textsuperscript{34} See supra note 6 and accompanying text.
\textsuperscript{35} SEGAL & SPAETH, supra note 7, at 204.
fixed attitudes about constitutional law generally or in particular fields. Third, researchers could identify which materials are pertinent for demonstrating nominees’ judicial attitudes and assess nominees’ attitudes on the basis of those materials. For instance, nominees’ speeches or law review articles might reflect certain ideological commitments. This might be especially pertinent for nominees to appellate courts, particularly the Supreme Court. Whereas district court judges are presumably bound by the opinions of the federal court of appeals within their respective circuit, appellate judges are bound by Supreme Court opinions, but less so, and sometimes not at all, by the opinions of other panels within their respective circuits. The extent to which judges may not be strictly tethered by precedent requires examining the various other bases on which they might ground their opinions. The problems with this approach are, however, that some nominees might not have produced any extrajudicial writings, some writings or speeches might not reveal much about judicial nominees’ attitudes, and the nominees can credibly argue that the speeches or articles are irrelevant because scholars and judges are not subject to the same constraints. Scholars are free to analyze legal questions on the basis of whatever sources they consider to be appropriate, while judges, even appellate ones, are not. The latter are duty-bound to assess legal questions on the basis of a relatively narrow range of materials, including precedent.

It is possible that measuring merit is complicated by yet one more factor: We have no consensus on whose opinion ought to matter in evaluating merit. Law professors, practitioners, and even other judges may have special insights into what jurists do, but there is good reason to think that none of these groups are perfectly neutral. Each group qualifies as an external source of authority on merit, but each group’s members may only employ the conceptions of merit in which they have vested interests, with each gravitating toward the judges and Justices with whose opinions they tend to agree.

None of the deficiencies of particular definitions of merit, however, make empirical studies of judicial performance and selection impossible. We need not settle on a normative conception of merit on which everyone in the world can agree. What we need are normative criteria that are at least coherent, credible, and comprehensive. The next Part examines the missing element in most empirical research on judicial selection that precludes it from being as comprehensive and precise as it ought to be.

III. THE BATTLE OVER THE MAINSTREAM IN CONSTITUTIONAL LAW

The reasons for the attraction, or dominance, of ideology in judicial selection are obvious. First, national political leaders care about
ideology because of the high stakes involved in judicial appointments. They understand that Article III judges enjoy life tenure and thus are immune from political retaliation against their decisions. Judicial opinions on constitutional law cannot be overturned through ordinary legislation but only through the extraordinary means of reversal by constitutional amendment or by a superior court (for federal district and appellate courts) or by the Supreme Court (for its own opinions). Consequently, national political leaders devote a good deal of time trying to ensure that the people appointed as judges and Justices will exercise power in ways that are satisfactory to them.

Second, national political leaders have almost no incentive to reach any consensus on merit. Most citizens pay little or no attention to lower court appointments, so leaders can expect little or no public backlash to their decisions on lower court appointments. Moreover, Presidents and Senators are reluctant to relinquish their institutional prerogatives in the selection process. If they ever do so, it is only in exchange for something else that they have decided is more important to them (at least for the moment). Presidents and Senators might sometimes have incentives to reach accommodations, but accommodations are much harder to come by for Presidents and Senators from the opposition party. Presidents from one party and Senators from the other often need conflict to sharpen the differences between them and to call attention to the stakes involved in the selection process. Bipartisan agreement on a general definition of merit would merely reduce, rather than preserve or expand, Senators’ discretion in subsequent confirmation proceedings. It would tie Senators’ hands in specific confirmation contests. Merit gets attention when there is a political advantage in addressing it.

The most intense confirmation contests between Republicans and Democrats focus not on merit but rather on the contours of the mainstream in constitutional law. Each side claims that its nominees are in the mainstream and that the other side’s contested nominees are outside it. For instance, Senators opposed to the Bork nomination argued that he was outside the mainstream of constitutional law, while his defenders argued that his scholarship and thinking was well within it. More recently, Democrats have supported six filibusters against judicial nominees whose views on constitutional issues are, in the Democrats’ judgment, outside of the mainstream. The defenders have argued that just the opposite is true.

The contest to define the mainstream has not just been rhetorical. A good deal is at stake. Each side desperately wants its nominees to be viewed as occupying the middle, rather than the extreme end or outside, of the spectrum in constitutional law. The middle is the safest, strongest ground. Moreover, opposing nominees, because they are outside the mainstream, puts the other side on the defensive. More important, each side appreciates the enormous stakes involved, for with each victory each side advances one step further in building a foundation for an enduring constitutional vision. The vision is important in guiding not just other judicial nominations but also the exercise of presidential and legislative authority. The prize is shaping constitutional law for as far into the future as possible.

While it is not hard to understand why political leaders care intensely about securing the mainstream—or the middle—in constitutional law, it is harder for someone outside of (or not invested in) the process to determine what counts as the middle. While this determination would be useful for analyzing the claims of the opposing sides in defining the mainstream, I consider in Part III.A some of the difficulties with determining the mainstream in constitutional law. With these difficulties in mind, I then propose ways in which we might figure out the mainstream, or the middle, in constitutional law.

A. Problems with Defining the Middle in Constitutional Law

There are several major problems with identifying the middle ground in contemporary constitutional law. First, empirical analysis cannot easily capture what counts as the middle because the choices of what to emphasize or count are value-laden. Anyone looking to define the middle, or the mainstream (and the two are not necessarily the same), in constitutional law must make judgments about relevance: Are all cases relevant? Should we only look at the judgments or outcomes in particular cases, or should we also look at the reasoning (including its quality and extent)? Where, for instance, do seemingly obvious cases like *Roe v. Wade*, 39 *Lawrence v. Texas*, 40 and *Lee v. Weisman* 41 fit? Some might argue that they are clearly on the “left” in constitutional law, but others might argue that they are consistent with a libertarian perspective on the “right.” Arguing that one or the other of these positions is correct is just another value judgment.

Second, an even more serious problem with defining the middle or the mainstream in constitutional law is that neither the categories we employ in assessing judicial performance nor many nominees’ constitutional views are fixed. Because of the phenomenon of ideo-

logical drift, categories are not static; particular perspectives on constitutional law associated with particular political factions may over time be appropriated by or become associated with different political factions. For instance, Chief Justice John Marshall reflected a “conservative” rather than a “liberal” perspective on constitutional law, because he usually favored the status quo. His successor as Chief Justice of the United States, Roger Taney, was understood, at the time of his appointment, as representing a “liberal” perspective on constitutional law because he was thought to favor progressive legislation and reform of the status quo. It is only because of ideological drift that each is now viewed differently. New Deal liberals found they had a lot in common with the opinions of Chief Justice Marshall, because of their consistent support for a strong national government; and conservatives admired Taney’s ardent efforts to resist the expansion of the national government at the expense of state sovereignty.

The labels “liberal” and “conservative” do not fit contemporary Justices much better. For instance, Justice John Paul Stevens, appointed to the Court in 1976 by President Ford, is frequently described as a “liberal” by commentators and critics. Yet he hardly seems to have much in common with other “liberals” such as Associate Justices William Brennan and Thurgood Marshall, with whom he sat for many years. Nor does it seem appropriate to describe Justices Kennedy and O’Connor as strictly “conservative” simply because they have often favored protecting state sovereignty in Commerce Clause and Eleventh Amendment cases. They also have voted to reaffirm the embattled decision of Roe v. Wade, to strike down antisodomy laws in Lawrence v. Texas, and to strike down the Virginia Military

---

44. See id.
49. 539 U.S. 558 (2003) (majority opinion of Kennedy, J., joined by Stevens, Souter, Ginsburg, and Breyer, JJ.); id. at 579-85 (O’Connor, J., concurring in the judgment).
Institute’s policy to exclude women.\textsuperscript{50} It is conceivably more accurate to describe Justices Kennedy and O’Connor as moderates, though many Senators would resist doing so because it would cede the middle ground to these Justices rather than to others they might prefer to place there.

A third problem with fixing the middle ground in constitutional law is that Justices and judges sometimes shift their attitudes about constitutional law either generally or in particular cases. Justice Harry Blackmun is often described as evolving, or growing, over time into a more “liberal” justice.\textsuperscript{51} Others might move in the other direction. Segal and Spaeth claim, inter alia, that Justices Stevens and Souter each became more “liberal” over time, while Justice White became more “conservative” over time.\textsuperscript{52} As Chief Justice, William Rehnquist has sometimes been said to have moderated some views,\textsuperscript{53} as he arguably did in writing the Court’s opinions reaffirming \textit{Miranda v. Arizona},\textsuperscript{54} concurring in \textit{United States v. Virginia},\textsuperscript{55} and upholding the Family Leave Act as an exercise of Congress’s authority pursuant to Section Five of the Fourteenth Amendment in \textit{Nevada Department of Human Resources v. Hibbs}.\textsuperscript{56}

Some people seem to assume that most nominees have ideological precommitments at the times of their nominations that are impervious to change, but it is impossible to prove that this is true, especially when the nominees themselves disclaim holding any such commitments.

Fourth, legal academics have done little to illuminate what may fall inside or outside the mainstream of constitutional law. Most legal scholars appear interested less in finding common ground than in delivering the knock-out punch against opposing points of view.\textsuperscript{57} A common goal of legal scholarship is paradigm-shifting, but in pursuing this goal, legal scholars will dismiss as wrong or dangerous points of view affiliated with the paradigm they are trying to undo. The pursuit of this goal is not likely to enrich our understanding of

\begin{itemize}
\item \textsuperscript{50} United States v. Virginia, 518 U.S. 515 (1996) (majority opinion of Ginsburg, J., joined by Stevens, O’Connor, Kennedy, Souter, and Breyer, Jd.).
\item \textsuperscript{51} See, e.g., Bernard Schwartz, \textit{A History of the Supreme Court} 316 (1993).
\item \textsuperscript{52} Segal & Spaeth, supra note 7, at 218.
\item \textsuperscript{53} A number of suggestions have been made as to why Chief Justice Rehnquist seems to have moderated some views. See, e.g., Yale Kamisar, \textit{Foreword: From Miranda to § 3501 to Dickerson to . . .}, 99 Mich. L. Rev. 879, 889-92 (2001).
\item \textsuperscript{55} 518 U.S. at 558-66 (Rehnquist, C.J., concurring in the judgment).
\item \textsuperscript{56} 538 U.S. 721 (2003).
\item \textsuperscript{57} Cf. Suzanna Sherry, \textit{Too Clever by Half: The Problem with Novelty in Constitutional Law}, 95 NW. U. L. Rev. 921 (2001) (noting that in many scholarly discussions of the counter-majoritarian difficulty, “grandtheory” tends to dominate over simpler and more coherent arguments).
\end{itemize}
which views actually do, rather than ought to, fall within the mainstream of American constitutional law.

Fifth, the media hinders sophisticated discussions of judicial performance. The media has begun to shirk its traditional role in educating the public. It has moved from reporting “hard” news, or facts and figures, to reporting “soft” news, or speculation and commentary. The proliferation of media outlets and twenty-four-hour news has put enormous pressure on newspapers and television reporters to emphasize scandal. The media prefers drama and conflict, because it gets people’s attention. As candidates and commentators increasingly feel the need to characterize opponents in extreme terms, the media follows suit. Candidates are thus “liberal” or “conservative,” and Justices are also one or the other. No one, apparently, begins as a moderate or ends up as one. From the perspective of the media, the middle in politics is nothing more than the otherwise unoccupied ground that the candidates fight to control, while the media simply covers the flashier portions of the fight.

B. Sketching the Middle

Assessing judicial ideologies is difficult without having some yardstick with which to measure them. One cannot talk about extreme views, or views falling outside of the mainstream, without clarifying which views are not extreme, or do not fall outside the mainstream. It is possible that the measurement of an ideology is a purely normative matter, depending on its appeal to lawmakers and its consistency with constitutional law as they understand it. Even then we need to define the middle, or moderation, as a means of curbing reckless or misleading rhetoric. We need our rhetoric to fit the complicated business of judging. So the question is how accurately can we describe a middle course or the contours of the mainstream in constitutional law.

I offer a few possible answers, with each of the difficulties described above in mind. First, we can identify the middle ground as that which each of the contending sides in confirmation contests is trying desperately to occupy. We can define it, in other words, as an aspiration. We can assess nominees based on how well they fit the description of the middle ground, or mainstream, of their supporters. One problem with this definition is that it might allow one side to define the terms on which it prefers for its nominees to be assessed, without any second-guessing; however, this understanding of the mainstream puts pressure on supporters of particular nominations to

58. See Hess, supra note 9.
be careful about how they characterize nominees or risk having the latter fail to meet expectations.

Second, we could define the mainstream as comprised of the pool of people who successfully made it through the judicial nomination process. They constitute a large and diverse pool, which reflects the approval of the governing elite. The problem with this understanding of the mainstream is that it fails to take into account the fact that many people make it through the process without close scrutiny. Many of the people confirmed also might not have had fixed views at the time of their appointments or might change their views over time. In addition, it is not clear why we should ever define the mainstream based on what judges and Justices actually do, because they act independent from the governing elite once they are confirmed. Indeed, most judges and Justices serve long after the political coalition or majorities who chose the people responsible for their appointments have ceased to exist. For instance, John Marshall served most of his long tenure as Chief Justice long after the demise of the Federalist party with which he had been associated.

Moreover, defining the mainstream as those whom the Senate has confirmed merely gives each side an incentive to push the envelope. With each victory in the confirmation process, each party has expanded the possibilities for its nominees. Once people are confirmed, their parties can point to them as examples, or precedents, to guide future confirmation proceedings.

A more interesting but speculative test might be to ask whether the President would still nominate or the Senate still approve the same judge if they knew what kinds of decisions the judges would make. In many cases, nominees are relatively blank slates, and, in any event, judges and Justices presumably fulfill special obligations independent from presidential and senatorial influence. So it might not be fair to attribute to Presidents and Senators all the decisions made by the judges and Justices they have approved.

Third, the mainstream could be understood as simply consisting of the views of those at the center of the Court. These days that would presumably be Justice O’Connor, because she rarely dissented in the 2003 Term.59 The problem is that she did not decide these cases alone, and it is unclear why those with whom she joined in majority opinions ought to be excluded from the mainstream. Moreover, the center can shift, and there is no guarantee that Justice O’Connor will be there as often next year. Nor is it clear why dissenters ought to be excluded entirely, because dissents sometimes later become the law.

The fourth and final possibility is to define the mainstream as something more dynamic and broader than a specific Court or specific Justice at a particular moment in time. The Court is not alone in making constitutional law. Our political leaders make a great deal of constitutional law, much of which eludes judicial review. Moreover, the Court approves the vast majority of the constitutional decisions that it does review. It would also be wrong to assume that every Supreme Court decision reflects mainstream constitutional values. Sometimes the Court gets it wrong, as it did in *Chisholm v. Georgia*, *Dred Scott v. Sandford*, and *Korematsu v. United States*. The constitutional views of Presidents and Senators are relevant to the makeup of the mainstream, because they have the power to try to move the Court in different directions (or perhaps keep it on course) by virtue of their respective authorities in the appointments process. They also have the power to shape the size, direction, and priorities of the federal government. Moreover, Presidents and members of Congress perform critical roles in approving enduring constitutional changes and in settling constitutional crises—failures within the Constitution to provide solutions to certain kinds of disputes. Consequently, it is possible to define the mainstream as the dominant doctrine, outlook, and thinking on constitutional law in a given period. The Court provides formal doctrine, the courts and national and state political leaders shape the constitutional outlook of a particular era, and all of these along with constitutional commentators and historians (in a wide variety of fora) provide critical thinking on constitutional law. This perspective on the mainstream has the virtue of encapsulating the constitutional activities of a given era. Its problem is that there is no method on which all people could agree for determining the relevant doctrine, outlook, and thinking of a particular era. Historians might be in the best position to pull this information together, but only in retrospective. It is a challenge, to say the least, for someone to step outside of his or her own time to develop a credible perspective on it. Time does not stand still for any person or any thing, and none of the things that we might think are essential for shaping the mainstream—doctrine, outlook, and critical thought—are purely static.

---

61. 2 U.S. (2 Dall.) 419 (1793).
62. 60 U.S. (19 How.) 393 (1856).
63. 323 U.S. 214 (1944).
IV. PROVING (HOW MUCH) IDEOLOGY MATTERS

Proving that ideology is a dominant factor in the judicial selection process—as many people suppose—is no small feat. There are a number of complications with determining the extent to which ideology was a major factor or the primary basis for the President’s nomination and the Senate’s confirmation of various judicial nominees. Below, I first review these problems and then offer some modest suggestions for future empirical analysis on the significance of ideology in the judicial selection process.

A. Fixing Ideology

There are a number of problems with proving empirically whether and, if so, how much ideology was a factor in the nominating or confirmation phase. To be sure, this is a question that social scientists have spent considerable time and effort trying to answer. Nevertheless, several problems persist. First, reaching consensus on what qualifies as ideology is difficult. While I personally understand ideology as a precommitment to certain constitutional values or to resolving particular questions of constitutional law, regardless of the facts of particular cases, this is but one understanding. Indeed, people widely disagree over how to define ideology and even whether Presidents and Senators have taken ideology into account in the appointments process. Consequently, one difficulty with proving that ideology matters in the selection process is adopting a credible definition of ideology.

The second problem is that it is a mistake to assume that every judicial nominee has a well-conceived or thoroughly worked-out constitutional ideology. It is possible that many, even most, do, but judicial nominees often publicly disavow commitment to a particular constitutional ideology. Moreover, ideology presumably functions as a blinding mechanism, so that it is conceivable that some nominees may not be aware that they have certain ideological commitments.

Third, and perhaps most important, the relevance of ideology to particular judges’ decisions or to particular confirmation decisions may not be evident in the public record. If the President and his nominees deny that the latter have particular ideological commitments, then the burden shifts to the other side to prove them wrong.

65. See, e.g., Charles M. Cameron et al., Senate Voting on Supreme Court Nominees: A Neoinstitutional Model, 84 AM. POL. SCI. REV. 525 (1990); Jeffrey Segal, Senate Confirmation of Supreme Court Justices: Partisan and Institutional Politics, 49 J. POL. 998 (1987); Jeffrey A. Segal et al., The Role of Ideology in Senate Confirmation of Supreme Court Justices, 77 KY. L.J. 485 (1988-89); Donald R. Songer, The Relevance of Policy Values for the Confirmation of Supreme Court Nominees, 13 LAW & SOC’Y REV. 927 (1979).

This is precisely the dynamic with President Bush’s judicial nominations. He has publicly defended his nominations on the ground of merit, and he has disavowed that he has employed a litmus test or chosen nominees based on particular ideological commitments. His nominees also publicly disavow such commitments. Consequently, skeptical Democrats must infer his selection criteria—including any preference for ideological precommitments—from the kinds of nominees that he has chosen.

Moreover, most judicial nominations do not fail because the Senate formally rejects them. They fail because of inaction. Senate rules provide that a nomination lapses unless the Senate has acted on it before the end of the current legislative session, and if they do not, the nominations fail. For instance, in President Clinton’s final year in office the Senate failed to act on more than sixty of his judicial nominations. There is little or no record on these nominees, so it is not possible to prove precisely why the Judiciary Committee did not hold hearings or votes on these nominees. The official record is silent on why these nominations failed.

To complicate matters further, the Senate debates that do occur over nominees rarely employ the term “ideology.” More often than not, the focus in confirmation contests has been on such matters as the nominee’s integrity, experience, competence, and temperament. When the debates do shift focus to nominees’ commitments to (or expression of) particular constitutional views, they feature discussions about whether the nominee comes from the “mainstream” of constitutional law.

Fourth, proving that ideology significantly matters to the fates of judicial nominations is complicated by the fact that Presidents and Senators rarely base their decisions in the appointments process on a single factor. Presidents, or their counselors, usually employ a range of criteria for making decisions on whom to nominate. In the Senate, a single factor is not necessarily determinative. It is possible that Senators might initially be disposed against particular nominations for one or two basic reasons, but their public opposition may not necessarily be predicated on these. Nor is it unusual for Presidents and Senators to base public decisions on factors they do not disclose.

71. See generally ABRAHAM, supra note 4.
It is not incumbent upon our national leaders to disclose all the grounds for their constitutional decisions.

Moreover, Presidents and Senators must make different kinds of decisions in the selection process. Because Presidents are responsible for choosing nominees, they can make decisions about which persons they think are best qualified or best fit their selection criteria. Senators use a different calculus. While they are often able to provide input (and even make specific recommendations) on nominations, their primary responsibility is to determine, not necessarily whether the nominee is ideal or the best qualified, but rather whether the nominee is acceptable according to whatever criteria each Senator decides is relevant. It is thus not unthinkable that 100 different Senators may use 100 different sets of criteria for evaluating judicial nominees.

B. How to Show Ideology Matters

The aforementioned problems are not necessarily fatal to the enterprise of proving that ideology makes a difference to outcomes in the confirmation process. Patterns invariably emerge within the process. For instance, the Senate usually (but not always) approves the vast majority of a President’s judicial nominations. One could thus try to identify what successful nominees have in common or what traits or characteristics are shared by unsuccessful nominees. These are not necessarily easy ventures, but they are not impossible. For instance, social scientists, including David Yalof, have shown what they regard as the characteristics that the people nominated to the Supreme Court over the past few decades have had in common.72

Once one sets out to demonstrate the particular significance of a single factor, such as ideology, the task becomes somewhat more complicated. To make this showing, one needs to first determine the relevant independent and dependent variables.73 The variable that is to be explained—in the case of the confirmation process, the vote share (or how Senators voted on particular nominations) or the absence of a vote on a particular nomination—is the dependent variable; it “depends,” or turns, on other variables. The latter are what social scientists call independent or “explanatory” variables, because they help to explain the dependent variable. The independent or explanatory variables are not themselves explained by the theory one is trying to prove; they simply do the explaining. The effects of these variables are called coefficients. Mapping the coefficients on graphs

72. DAVID ALISTAIR YALOF, PURSUIT OF JUSTICES: PRESIDENTIAL POLITICS AND THE SELECTION OF SUPREME COURT NOMINEES (1999); see also Epstein et al., supra note 1.

73. For an excellent primer on statistical analysis, see RAY C. FAIR, PREDICTING PRESIDENTIAL ELECTIONS AND OTHER THINGS (2002).
(to determine the points at which the variables intercept) and testing the coefficients are relatively sophisticated steps, which must be done in statistical analysis. But my point here is to clarify that, as an empirical matter, ideology is not the outcome that needs to be proved; instead, it is one of numerous variables that determine outcomes in the confirmation process. Thus, we need to determine the relative impacts of ideology and other independent variables in the confirmation process.

1. **Proving Judicial Ideology**

One cannot prove that ideology matters without initially determining how ideology manifests itself. Presumably, it must reflect some commitment to a particular approach to constitutional interpretation, regardless of the facts of particular cases. So the challenge for an empiricist is to track down and assess the intensity of a nominee’s expressions of any such ideological commitments. As I have suggested, a common approach among social scientists is to find these expressions by means of external indicators such as newspaper editorials and sponsoring Senators, even though these indicators may not all be reliable or credible. These indicators are by no means the only ones. Indeed, I suggest almost a dozen possible indicators of nominees’ ideological commitments, each of which can be empirically measured.

First, if the nominees are judges, one can inspect their opinions. Opinions merit special attention because judges speak through them. They have enormous potential to reveal what, if any, ideological commitments are at work. If opinions contain judges’ explicit acknowledgments of ideological commitments, they ought to be given special weight, particularly because judges are not obliged to make such acknowledgments in constitutional adjudication. If opinions instead reflect patterns of decisionmaking, then these too merit special consideration. This is especially true if the patterns relate to methodology or prioritization of sources of decision.

Second, nominees’ writings and speeches might illuminate their ideological commitments. These materials are important, to some degree, because some people are nominated in part because of the views expressed within them. Their writings and speeches are important, in other words, because they attracted attention from the right people. If nominees’ writings and speeches contain the nominees’ admissions of certain ideological commitments, they merit special weight. If they do not, then patterns in methodology or prioritization of sources could become relevant, especially if they are nearly or absolutely uniform.
Alternatively, one could merely award a number to a nominee, which would be based on the number of his or her extrajudicial writings on controversial subjects. The actual content of these writings may be less important than the number, because it is possible (and thus one would need to show) that the larger the number of writings on controversial subjects, the more a nominee has opened himself or herself to attack for his or her views on those subjects. (It is, of course, also possible that nominees can make controversial statements about relatively mundane topics.) Thus, especially prolific constitutional scholars would likely receive very high numbers, which could be negatively weighted because of the likelihood that some Senators will try to use their writings against them.

Fourth, other possible indicators of ideological commitments may be nominees’ professional activities. Some activities are more revealing than others. For example, someone who has dedicated a great deal of time to opposing the death penalty as unconstitutional is likely to have a settled view on the matter. It is, however, possible for some nominees to argue that their professional activities were merely undertaken at the bidding of their superiors or clients. Thus an empiricist needs to carefully screen nominees’ professional activities. The more they fall into controversial areas, the more negative weight they can be given in the confirmation process.

Fifth, one might ask which groups support particular nominees and on what bases. Groups no doubt might have different reasons to support different nominees, but some groups may be well known for preferring nominees with certain kinds of ideological commitments. The preferences of the latter group might be particularly relevant, especially if the group has identified the grounds for its approval of the nomination.

Sixth, one could measure which groups are opposing which nominees and the grounds for their opposition. Of particular interest will be the ones charging unacceptable ideological commitments. The more groups that do this, the more likely Senators will know about their opinion.

Alternatively, one could simply measure the numbers of witnesses testifying for and against particular nominations. While this data is obviously not available for nominees who never receive hearings, it could be very useful for illuminating the dynamic within particular confirmation contests. The larger the number of witnesses testifying, the more likely it is to be significant. There must be a reason why one nomination draws more witnesses than another. That reason may or may not have anything to do with ideological commitments, but it does bear on the nominee’s likely fate. So an empiricist needs to determine the grounds on which witnesses testify for and against
particular nominations, which can then be factored into his or her analysis.

Seventh, one could consult an administration’s selection criteria. Through their nominating authority, Presidents are able to set the terms of debate in the confirmation process. If nominees are not presented based on their ideologies, the burden shifts to the other side, and in practice it is hard for Senators on the defensive in the confirmation process to direct its results.

Eighth, empiricists ought to consider measuring the amount of time devoted in hearings to discussing nominees’ ideologies. This requires counting not only the number of Senators’ statements and questions pertaining to ideology but also how much time nominees spend defending their judicial philosophies or explaining their supposed ideologies. It is hard to dismiss a large amount of attention as nothing more than pretext. Even if it were, it shows the particular role played by ideology in particular hearings.

A related empirical inquiry could be undertaken to simply measure how much time nominees spend defending themselves against charges in their confirmation hearings. A rule of thumb in confirmation hearings is that the more time nominees spend testifying before the Judiciary Committee (the “Committee”) the worse their chances of confirmation become. This is not because they are bad witnesses, though they may be. The concern, however, is that the longer the nominee is required to testify, the more likely there is something problematic about the nominee. Once these figures are determined, they could be grouped according to the reasons for extended testimony.

Ninth, people who are not judges, even if they are academics, can credibly claim that their public musings do not reflect what they would do as judges because their duties as judges would require them to abide by certain norms, such as following precedent, that they are not bound to as scholars or commentators. These disclaimers are made under oath and thus need to be factored into empirical analysis as well, for they shift the burden to the opposition not just to disprove them but also to suggest the nominees lied under oath.

Tenth, a final measure of ideological commitments may come from the testimony of those who claim to know the nominees best. What, in other words, do the people who claim to have read the nominees’ writings or opinions claim about the nominees’ ideological commitments? This may not be entirely reliable, because many of the people contacting or testifying before the Committee may have vested interests in the outcomes of the hearings. To avoid the latter problem, it might be useful to examine whether these people are testifying against their nominal parties’ interests. For instance, Michael
McConnell benefited from having a number of prominent Democrats urge the Committee to approve his nomination, even though some Democratic Senators were disposed to be against it.\textsuperscript{74} While McConnell had produced a relatively high number of extrajudicial writings on quite controversial subjects, that number could be reduced, or counterbalanced, by a different number based on the size of the group urging his confirmation against their own party’s supposed interest.

Last but not least, empiricists ought to consider examining all failed nominations in order to determine what traits, if any, they had in common. It is possible that they all, or most of them, attracted opposition based explicitly on ideological concerns. But they also might have attracted opposition because of a confluence of several of the other factors that are discussed \textit{infra} in Parts IV.B.2 through IV.B.10.

2. \textit{Senate Composition}

Confirmation proceedings do not occur in a vacuum. Among the other factors likely to have an impact on their outcomes is the Senate’s composition. Presidents often take the composition of the Senate into account in deciding whom to nominate and when. The representative strength of a President’s political party in the Senate is obviously important, because it determines which parties control the Judiciary Committee, the agenda on the floor of the Senate, and the length of debate. If the minority party controls at least forty seats in the Senate, it can then block some judicial nominations by filibustering them.\textsuperscript{75} Threatening filibusters or temporary holds (which are, in effect, mini-filibusters) can sometimes influence whether and when Presidents make certain nominations.

The Senate composition can thus influence how quickly and even how many nominations get through to the Senate floor for final votes. If the opposing party controls the Senate, it raises the likelihood of obstruction of at least some judicial nominations. Hence if the President’s party does not control the Senate, this number is likely to be a negative coefficient. If, however, the President’s party does control the Senate, it is likely to be a positive one, and quite large if the President’s party controls more than sixty seats. If the President’s party controls a majority but less than sixty seats, it still merits a positive coefficient, though the possibility of filibuster merits making it not a very high one.


3. Timing

Another factor that can potentially influence confirmation outcomes is timing. Election years tend not to be good times for Presidents to make judicial nominations, particularly to the Supreme Court. In the nineteenth century, the Senate did not act on at least nine Supreme Court nominations, supposedly because the majority party was trying to keep the vacancies open until after the next presidential election. In President Clinton’s final year in office, the Senate did not act on more than sixty of his judicial nominations. Similarly, the Democratic-led Senate did not act on dozens of the first President Bush’s judicial nominations in his final year in office, presumably because of a desire to reserve as many judicial vacancies as possible for the next President.

Timing might matter in a different way. Most failed nominations do not get so far as receiving Committee votes; they fail, as I have suggested, because of inaction. Moreover, not all nominations that get hearings are scheduled for Committee votes. Consequently, one needs to figure out how long a nomination has gone without a hearing or whether it has gotten a Committee vote within a certain period of time (presumably the average length of a time between a nomination and a Committee vote). Those nominations exceeding the average length of time without yet getting a hearing or Committee vote will likely not be approved. Moreover, the closer a nomination comes to being made near the end of a legislative session, the less delay is needed to nullify it.

Consequently, timing can be quantified in at least three ways, each negatively. The first is the extent to which it exceeds the average length of time for confirmation; the second is whether the nominations have been made in an election year; and the third is the extent to which a nomination was made with less time left in a session than the average amount of time needed for confirmation.

4. Sponsoring Senators

Sponsoring Senators may make a difference to the fates of at least some judicial nominees. The more powerful the Senator, the more likely one might expect nominees he has supported to be confirmed. For instance, the Senate has confirmed almost a half-dozen nominees who at one time or another worked for Senator Orrin Hatch, the former Chair of the Judiciary Committee. Indeed, Senator Hatch convinced President Clinton to nominate a former aide to a district court

76. See ABRAHAM, supra note 4.
77. See Hardin, supra note 69.
78. See Goldman, supra note 1.
in Utah, after he had held up every other judicial nomination pending President Clinton’s compliance.80

5. Presidential Popularity

It is possible that a President’s popularity might have an effect on a nomination’s fate. By popularity, I mean the President’s political strength as reflected in, inter alia, his approval ratings with the public. These might show the risks involved in a fight with the President. The more closely a nomination is identified with the President (or the more it means to him) or his policies, the more likely that the popularity of the President or his policies will be an important factor. The more popular the President or the policies with which the nominee is associated, the greater likelihood this popularity will benefit his nomination. The more political coinage that a President has on which to draw from in confirmation contests, the more likely Senators will suffer some political damage or loss from such confrontations. Some Senators might choose contests over some judicial nominations because they believe the conflicts can improve their standing with important constituencies or can underscore their own political commitments. But contests are not likely to be completely cost-free, particularly insofar as Presidents remember them and have the means and opportunity to seek retaliation.

A related factor may be party cohesion or fidelity. The extent to which Senators from the same party are willing to stand together on judicial nominations makes a big difference as to whether they can successfully filibuster or defeat nominations in Committee or on the Senate floor. The degree of cohesion or unity within a caucus is pertinent to how much power it can wield under the Senate rules and its influence in striking deals with the President. Sometimes Senators do not do what their party leaders or Presidents from their parties tell them. Sometimes divisions in the ranks of the Senators from the President’s party are a problem for many nominees, with some joining members of the opposition party to defeat them.

6. The Blue-Slip Process

Another factor with potential influence on the confirmation process is whether the (and which) blue-slip process is in place at the time of a nomination.81 The blue-slip process allows a Senator to block a nomination made to an office in that Senator’s home state. This process is usually available to Senators from both parties, but sometimes Presidents or Senate leaders have restricted it to Sena-

80. See Gerhardt, supra note 33, at 307.
tors only from the President’s party. If this process is in place in either form, it expands Senators’ opportunities to block nominations. It particularly reinforces the strength of the majority party in the Senate. If that party is targeting the expression of support for particular policies or ideologies, then nominees who can be shown to have made such expressions face potentially serious obstacles to their confirmation from the outset.

The blue-slip process for particular nominations can thus be shown in at least two ways. First, it can simply be shown as the number of times it has been invoked to block nominations. This number reflects its vitality at a given moment in the confirmation process. Alternatively, and perhaps more meaningfully, an empiricist can measure what, if any, characteristics nominations blocked through the blue-slip process had in common. The number of nominations with supposed ideological commitments that triggered senatorial concern can then be calculated. It will be a fraction of the total number blocked through the blue-slip process.

7. **Numbers of Witnesses**

The numbers of witnesses called for and against nominees are likely to be pertinent to their chances of success in the confirmation process. The number of people testifying, particularly against a nominee, is likely to signal some problem with the nomination. If more people are testifying against a nominee than for him or her, the nomination is almost certainly in trouble. Of course, these numbers alone do not indicate the reasons for support or opposition. Some people may be opposed because of the nominee’s supposed ideological commitments, but one must go behind these numbers in order to determine this information.

8. **The American Bar Association**

The American Bar Association’s ratings on nominees may affect the fates of nominations. Positive ratings do not guarantee confirmation, but negative or largely unfavorable ratings are bound to considerably lower a nominee’s chances for confirmation. Even split ratings can be a problem (though not always fatal). The American Bar Association comes as close as any group to providing a “neutral” assessment of a nominee’s qualifications, and its ratings may be used by either side in a confirmation contest depending on the extent to which they are favorable or unfavorable. Thus, its assessment needs to be factored into any analysis of the process.
9. Integrity

Nominees ought to get an integrity rating. We expect judicial nominees to be honest people, whose integrity is beyond reproach. More than a few judicial nominations have failed because of nominees’ ethical lapses. One problem (of many) torpedoing Clement Haynsworth’s nomination to the Supreme Court was his participation in some cases involving companies in which his wife had owned stock. Less than two decades later, Douglas Ginsburg asked President Reagan to withdraw his nomination as an Associate Justice because he had failed to inform the FBI during background checks that he had smoked marijuana while he was a tenured Harvard Law School professor.

10. Merit

Of course, merit is a factor that cannot be ignored in evaluating the judicial selection process. The absence of merit, or questions about nominees’ qualifications, can be fatal to judicial (and of course many other) nominations. Even though we lack consensus on merit, it would be silly to dismiss it as irrelevant to the outcome of confirmation proceedings. In an empirical analysis of the selection process, we must, at the very least, measure the extent to which witnesses before the Judiciary Committee and members of the Committee focus on merit in their statements or questions. Presumably, the more it expressly comes up in hearings, the more potential impact it may have on a nomination’s fate. Of course, merit might be a red herring or a pretext. It is thus very important to give special attention to what those who come closest to being neutral observers or less-self-interested parties have said about the merits of particular nominations. So we might give special weight to evaluations by the American Bar Association, or perhaps to leading scholars, on the merits of particular nominations.

Because so many people (particularly Republicans) have questioned the neutrality of the American Bar Association and of law professors generally, one might have to look elsewhere for neutral or less-self-interested evaluators of merit. These are just nine factors, besides ideology, that are likely to affect the fates of judicial nominations. The odds are that judicial nominations will not founder simply because of one of these factors. Moreover, it is possible, if not likely, that the stated grounds of opposition to judicial nominations might not be entirely credible; they might reflect, at least to some extent, a pretext to oppose a nomination. For instance, the Judiciary Commit-

82. YALOF, supra note 72, at 104-08.
83. BRONNER, supra note 37, at 332-35.
tee never acted on President Clinton’s nomination of Elena Kagan to the U.S. Court of Appeals for the District of Columbia in 2000. She never got a hearing, much less a vote, on her nomination, in spite of her strong credentials. No one expressed opposition to her because of her ideology. Instead, opposition, to the extent it was ever manifest in public, focused more on whether the appellate court to which she had been nominated had a caseload to justify filling all of the seats to which the President had nominated people. A4 Some people might view this opposition as merely a pretext to preclude the confirmation of someone whom the opposition party feared might be a liberal activist or who would then occupy a seat that it would have preferred for one of its own to occupy. After President Bush took office, Republican leaders acknowledged the court’s caseload justified filling all of its seats after all. A5 And President Bush then nominated Miguel Estrada to one of them. It is possible that at least some opposition to the Estrada nomination derived in part from a desire for payback, though the grounds cited related to Estrada’s temperament and possible judicial ideology. While payback is another possible factor that needs to be monitored in the confirmation process, it is hard to verify, because Senators rarely (but sometimes do) acknowledge that it is the basis for their opposition.

In the final analysis, proving that ideology significantly affects the fates of nominations is not easy. Proving it may be so difficult that many people simply opt for anecdotal evidence or opt for merely analyzing the appeal of a particular nominee’s ideology. After all, it is not necessary to prove that all nominees shared commitments to problematic ideologies, but rather only the ones that Senators end up choosing to oppose for stated or unstated reasons. The higher, or more powerful, the court to which someone has been nominated, the more likely Senators will be concerned about the person’s likely judicial ideology. It is thus likely that Senators will be more concerned with the likely ideologies of circuit court or Supreme Court nominees than district court nominees. In any event, as long as Senators do not fear the President, Senators remain relatively free to pick and choose which nominees to oppose and on what bases.

V. CONCLUSION

I close with a challenge. I challenge others to talk more openly about merit in judicial selection and particularly whether merit can be defined separately from ideology. If so, then we have to wonder

---

85. See Goldstein & Tucker, supra note 84.
why more scholars, Presidents, and Senators do not do so. If not, then we need to explain why our apparent failure to separate merit from ideology ought not to lead us to simply join forces with the social scientists who believe that judges are nothing more than policy-makers who just happen to wear robes.
FEDERAL JUDGES AND THE HEISMAN TROPHY

STEVEN GOLDBERG*

Choi and Gulati propose selecting as Supreme Court Justices those federal circuit court judges who are outstanding at their job. They note that “differences exist” between “the job of a circuit judge and that of a Supreme Court Justice,” but they ask whether “the current system” of selecting Justices does a better job than their proposal, and they conclude, “We do not see how it could.”

They are mistaken. If we look at highly successful Supreme Court Justices of the past, we find only one with lower federal court experience of any kind. Many distinguished federal circuit judges served with no distinction on the Supreme Court. It is those who served as governors, Senators, Attorneys General, judges on the highest state courts, and in private practice who show up on the lists of great Justices. I will provide the basis for this assertion below, and I will also point out the assertion is entirely unoriginal: numerous scholars have observed that prior federal judicial service is not correlated with success as a Justice.

Of course, this does not prove that Choi and Gulati are wrong. Perhaps if they had been around generations ago, they would have identified federal judges who would have turned out to be great Justices, or perhaps federal judges today are more talented than their predecessors. Moreover, the information I am going to provide is essentially anecdotal. Not only is there wide disagreement about who was, or is, a great Justice, there is a small database, the jobs of judges and Justices, as well as their reputations, are always evolving, and the opportunity for any kind of controlled experiment is lacking.

Nonetheless, it is pretty clear that Choi and Gulati are looking in a strange place if they are trying to find outstanding Justices. Given the historical record that I will survey, it is hard to explain why they have chosen to analyze the decisions of federal circuit judges. At times, they hint at one possible explanation: they are trying to give life-tenured lower judges an incentive “to work harder at their jobs.” Yet, not many people would trade harder-working lower court judges for a less capable Supreme Court. And, most of the time, Choi and

* Professor of Law, Georgetown University Law Center. I would like to thank Abigail Scaffe for her outstanding research assistance.

2. Id. at 31 n.11.
Gulati frankly say that they are trying to assemble a better Supreme Court.

Perhaps Choi and Gulati are counting something as important just because it is countable. It is harder to compare quantitatively the qualifications of an Attorney General and a Senator than it is to compare quantitatively two lower federal court judges. Even looking at the pool of state high court judges (who have a hint of the Supreme Court’s power in that they have the final say on state law) may have been too difficult, since state courts do not cite each other in any uniform way.

But whatever the reason, Choi and Gulati are spending a good deal of effort in an unpromising area. Here is the historical record.

Prior to the current nine, ninety-nine Justices had served on the Supreme Court. Twenty-four of these Justices had previously served on lower federal courts, including such respected federal circuit court judges as William Woods, Howell Jackson, Willis Van Devanter, and Fred Vinson.

4. Id. at 304; Choi & Gulati, supra note 1, at 26-32.
5. See TIMOTHY L. HALL, SUPREME COURT JUSTICES: A BIOGRAPHICAL DICTIONARY vi (2001). I exclude from analysis the currently sitting Justices, since one cannot fairly evaluate a Justice's career until it has ended.
7. Prior to his appointment to the Supreme Court, Justice Woods had served for two years on a state chancery court and for eleven years on the Court of Appeals for the Fifth Circuit. HALL, supra note 5, at 179. During his service on the Fifth Circuit, Woods earned a reputation as a “fair and hardworking judge.” Id.
8. In addition to brief service on lower state courts, Justice Jackson had served for seven years on the Court of Appeals for the Sixth Circuit, where he was known for his “diligence and patience.” ILLUSTRATED BIOGRAPHIES, supra note 6, at 268.
9. Justice Van Devanter had served shortly on a state high court and for seven years on the Court of Appeals for the Eighth Circuit. Id. at 313-14. When being considered for elevation to the Supreme Court, Van Devanter was recognized for his vast judicial experience and understanding of complex issues. See id.
10. Justice Vinson had served on the United States Court of Appeals for the District of Columbia for six years, during which time most of his decisions were upheld by the Supreme Court. Id. at 423. Vinson, who was known as “powerful and resolute,” “was effective enough as judge in a circuit court of appeals to have President Roosevelt select him to serve at the same time as head of an emergency court to decide OPA appeals.” SIDNEY H. ASCH, THE SUPREME COURT AND ITS GREAT JUSTICES 194 (1971).
Which of the ninety-nine Justices were highly successful? I looked at eleven leading lists that have been compiled over the years. Some are based on surveys of experts, while others reflect only the views of the scholar compiling the list. In an effort to reduce the biases, ideological and otherwise, that inevitably accompany the creation of lists like these, I compiled my own list of those Justices ranked as “great” on at least two of the prior lists.

It turns out that twenty-three Justices show up on my composite list of the greatest to have served on the Supreme Court. These

---


12. See Blaustein & Mersky, supra note 6; Great Justices, supra note 11.

13. See Asch, supra note 10; Frank, supra note 11; Hughes, supra note 11; Leahy, supra note 11; Lerner, supra note 11; Currie, supra note 11; Frankfurter, supra note 11; Nagel, supra note 11; Schwartz, supra note 11.


15. The following list identifies each of the twenty-three Justices as well as the sources that ranked each Justice as “great”:

**John Marshall**: Asch, supra note 10, at 25-39; Blaustein & Mersky, supra note 6, at 37; Frank, supra note 11, at 43; Hughes, supra note 11, at 58; Lerner, supra note 11, at 89-108; Norman W. Provizer, John Marshall and the Foundations of Judicial Power, in Great Justices, supra note 11, at 33-48; Currie, supra note 11, at 4-7; Frankfurter, supra note 11, at 783; Nagel, supra note 11, at 957; Schwartz, supra note 11, at 94-98.

**William Johnson**: Frank, supra note 11, at 43; Currie, supra note 11, at 7-9; Frankfurter, supra note 11, at 783; Nagel, supra note 11, at 957.

**Joseph Story**: Blaustein & Mersky, supra note 6, at 37; Frank, supra note 11, at 43; Hughes, supra note 11, at 58; F. Thornton Miller, Joseph Story's Uniform, Rational Law, in Great Justices, supra note 11, at 49-72; Currie, supra note 11, at 9-11; Frankfurter, supra note 11, at 783; Nagel, supra note 11, at 957; Schwartz, supra note 11, at 98-102.

**Roger Taney**: Asch, supra note 10, at 40-52; Blaustein & Mersky, supra note 6, at 37; Frank, supra note 11, at 43; Kenneth M. Holland, Roger B. Taney: A Great Chief Justice?, in Great Justices, supra note 11, at 73-109; Currie, supra note 11, at 11-13; Frankfurter, supra note 11, at 783; Nagel, supra note 11, at 957; Schwartz, supra note 11, at 102-08.

**Benjamin Curtis**: Frank, supra note 11, at 43; Hughes, supra note 11, at 58; Frankfurter, supra note 11, at 783; Nagel, supra note 11, at 957.

**John Campbell**: Frank, supra note 11, at 43; Frankfurter, supra note 11, at 783; Nagel, supra note 11, at 957.

**Samuel Miller**: Asch, supra note 10, at 55-70; Frank, supra note 11, at 43; Hughes, supra note 11, at 58; Currie, supra note 11, at 13-16; Frankfurter, supra note 11, at 783; Nagel, supra note 11, at 957.
twenty-three Justices cover a large swath of American judicial history, from John Marshall, Joseph Story, and Roger Taney to Robert Jackson, Earl Warren, and William J. Brennan, Jr. Only one of these

Stephen Field: Frank, supra note 11, at 43-44; Hughes, supra note 11, at 58; Frankfurter, supra note 11, at 783; Nagel, supra note 11, at 957; Schwartz, supra note 11, at 108-15.

Joseph Bradley: Frank, supra note 11, at 43; Hughes, supra note 11, at 58; Currie, supra note 11, at 16-18; Frankfurter, supra note 11, at 783; Nagel, supra note 11, at 957.

John Harlan (the elder): Asch, supra note 10, at 71-82; Blaustein & Mersky, supra note 6, at 37; Frank, supra note 11, at 43; Leahy, supra note 11, at 5-40; Linda C.A. Przylucki, The Evolution of John Marshall Harlan the Elder, in Great Justices, supra note 11, at 101-23; Nagel, supra note 11, at 957.

David Brewer: Frank, supra note 11, at 43-44; Hughes, supra note 11, at 58; Nagel, supra note 11, at 957.

Oliver Wendell Holmes: Asch, supra note 10, at 83-99; Blaustein & Mersky, supra note 6, at 37; Frank, supra note 11, at 43; Lerner, supra note 11, at 109-31; Patrick Garry, Oliver Wendell Holmes and the Democratic Foundations of the First Amendment, in Great Justices, supra note 11, at 125-45; Currie, supra note 11, at 19-23; Frankfurter, supra note 11, at 783; Nagel, supra note 11, at 957; Schwartz, supra note 11, at 115-22.

William Moody: Frank, supra note 11, at 43; Frankfurter, supra note 11, at 783; Nagel, supra note 11, at 957.

Charles Evans Hughes: Asch, supra note 10, at 117-27; Blaustein & Mersky, supra note 6, at 37; Frank, supra note 11, at 43; Leahy, supra note 11, at 41-72; Roger W. Corley, How Liberal Was Chief Justice Hughes?, in Great Justices, supra note 11, at 167-84; Currie, supra note 11, at 27-31; Frankfurter, supra note 11, at 783; Nagel, supra note 11, at 957; Schwartz, supra note 11, at 126-32.

Louis Brandeis: Asch, supra note 10, at 103-16; Blaustein & Mersky, supra note 6, at 37; Frank, supra note 11, at 43; Leahy, supra note 11, at 73-105; Lerner, supra note 11, at 132-48; Marguerite R. Plummer, Louis D. Brandeis: Pioneer Progressive, in Great Justices, supra note 11, at 147-55; Currie, supra note 11, at 23-27; Frankfurter, supra note 11, at 783; Nagel, supra note 11, at 957; Schwartz, supra note 11, at 122-26.

Harlan Stone: Asch, supra note 10, at 128-39; Blaustein & Mersky, supra note 6, at 37; Frank, supra note 11, at 43; Nagel, supra note 11, at 957.

Benjamin Cardozo: Asch, supra note 10, at 143-59; Blaustein & Mersky, supra note 6, at 37; Frank, supra note 11, at 43-44; Marguerite R. Plummer, Benjamin Cardozo: Pathfinder for Progress, in Great Justices, supra note 11, at 157-66; Frankfurter, supra note 11, at 783; Nagel, supra note 11, at 957.

Hugo Black: Asch, supra note 10, at 189-203; Blaustein & Mersky, supra note 6, at 37; Leahy, supra note 11, at 107-46; Lerner, supra note 11, at 149-60; Henry J. Abraham, Hugo L. Black as a Great Justice, in Great Justices, supra note 11, at 249-59; Nagel, supra note 11, at 957 & n.4; Schwartz, supra note 11, at 132-36.

Felix Frankfurter: Asch, supra note 10, at 160-71; Blaustein & Mersky, supra note 6, at 37; Lerner, supra note 11, at 161-66; Pederson & Provizer, supra note 11, at x; Nagel, supra note 11, at 957 & n.4.

William Douglas: Asch, supra note 10, at 204-17; Leahy, supra note 11, at 147-86; Lerner, supra note 11, at 172-76; Nagel, supra note 11, at 957 & n.4.

Robert Jackson: Asch, supra note 10, at 172-85; Lerner, supra note 11, at 167-71; Nagel, supra note 11, at 957 & n.4.

Earl Warren: Asch, supra note 10, at 218-28; Blaustein & Mersky, supra note 6, at 37; Leahy, supra note 11, at 217-58; Norman W. Provizer & Joseph D. Vigil, The Earl of Justice: Warren’s Vision for America, in Great Justices, supra note 11, at 261-78; Nagel, supra note 11, at 957 & n.4; Schwartz, supra note 11, at 136-42.

William J. Brennan, Jr.: Leahy, supra note 11, at 259-313; Rodney A. Grunes, Justice William J. Brennan, Jr., in Great Justices, supra note 11, at 279-302; Schwartz, supra note 11, at 142-50.
Justices—David Brewer, the author of \textit{Church of the Holy Trinity v. United States}—had ever served on a lower federal court.\footnote{16} I am not going to attempt to build a statistical argument on a small sample that turns on a vague term like “great.” But, when you have ninety-nine Justices, twenty-four of whom were lower federal court judges, you might expect that the twenty-four would be reasonably well-represented in a long, composite list of great Justices. When you find only one of them on such a list, you begin to suspect that lower federal court judges are not the best group to study when you are trying to identify successful Justices.\footnote{18}

The paucity of federal judges who were highly successful on the Supreme Court would not surprise anyone who has examined this matter. Indeed, many commentators have observed that prior judicial service of any kind is not a good predictor of Supreme Court success.\footnote{19} It is easy to pile up the quotations: “[H]istory proves that the best Supreme Court Justices are likely to be those who have not been judges before”;\footnote{20} “Oblivion has overtaken almost all” of the Justices who had served on lower courts;\footnote{21} scholars have found “a generally negative correlation between a justice’s ranking and prior judicial service.”\footnote{22}

This is not the place to examine why lower federal court service does not appear to prepare one for great success on the Supreme Court. Obviously, the jobs are different in many ways: having the fi-

\footnote{16. 143 U.S. 457 (1892). \textit{Church of the Holy Trinity} is famous for its approach to statutory construction and for its reference to the United States as “a Christian nation.” \textit{Id.} at 471. Brewer, the son of Congregational missionary parents, believed Christianity had been a great influence on American history, but he also believed American government was independent of any religion. STEVEN GOLDBERG, SEDUCED BY SCIENCE: HOW AMERICAN RELIGION HAS LOST ITS WAY 115-16 (1999).

17. Justice Brewer, who served five years on the Eighth Circuit Court of Appeals, also spent eighteen years on state courts where he issued two pioneering decisions that expanded the rights of women. ILLUSTRATED BIOGRAPHIES, supra note 6, at 252.

18. Moreover, Ross debunks the notion that Justices are viewed as “great” because they had prominent jobs like Attorney General before serving on the Supreme Court. Ross, supra note 14, at 418 (“With the exception of Cardozo and possibly Hughes, however, it is unlikely that the pre-Court careers of the ‘great’ justices influenced their rankings.”).

19. However, service on the highest court in a state has certainly produced some well-regarded Justices. Justice Holmes, for example, appears on nine lists of “great” Justices, see supra note 15, and served for twenty years on state courts, HALL, supra note 5, at 232. Consider as well Justice Cardozo, who appears on six lists of “great” justices, see supra note 15, and served for eighteen years on state courts, HALL, supra note 5, at 300.

20. FRANK, supra note 11, at 43.

21. Frankfurter, supra note 11, at 788.

22. Ross, supra note 14, at 419 (citing Thomas G. Walker & William E. Hulbary, Selection of Capable Justices: Factors to Consider, in BLAUSTEIN & MERSKY, supra note 6, at 52, 66). Another scholar who analyzed the results of the 1992 Pederson and Provizer survey concluded that prior judicial experience “is no guarantee of great success on the Supreme Court.” ROBERT C. BRADLEY, \textit{Who Are the Great Justices and What Criteria Did They Meet?}, in GREAT JUSTICES, supra note 11, at 1, 9.
nal say, always sitting en banc, the certiorari process, the political nature of many Supreme Court decisions, and many other variables may come into play.

One reason why we cannot be confident about the reasons lower federal judges have not been notably successful on the Supreme Court is that we know relatively little about the ingredients of such success. These ingredients are hard to pin down, and they are not the object of intensive study. More has been written about the parol evidence rule than about what makes a great Supreme Court Justice. Fortunately, there is an analogous area that has been the subject of broad public concern and careful analysis. That area is professional football, and the analogy suggests that Choi and Gulati are, indeed, barking up the wrong tree.

Applied to professional football, the Choi and Gulati thesis suggests that if you find the best college football player you will find a highly successful professional player. This thesis is far more likely to be true for football than it is for the judiciary. With the judiciary, Choi and Gulati have to contend with all of those other prominent people who were not federal judges but might become great Justices. Furthermore, they have to contend with a Supreme Court that plays by very different rules than the lower federal courts: it sits in groups of nine rather than three, it has certiorari power, and so on. No such problems exist in the football world. College football provides essentially all of the players in the National Football League, and college teams play by roughly the same rules as professional teams: each put eleven men on the field, have four downs to gain ten yards, and so on.

23. “We are not final because we are infallible, but we are infallible only because we are final.” Brown v. Allen, 344 U.S. 443, 540 (1953) (Jackson, J., concurring in the result).

24. The differences may, in fact, be subtler than are usually supposed. For example, Judge Harry T. Edwards of the U.S. Court of Appeals for the D.C. Circuit departs from many scholars in stressing the importance of collegiality in a federal circuit court, yet Edwards notes that “collegiality on the [Supreme] Court operates very differently from the collegial process at work in the lower appellate courts.” Harry T. Edwards, The Effects of Collegiality on Judicial Decision Making, 151 U. PA. L. Rev. 1639, 1644 (2003). Edwards points out that the Supreme Court hears many more “very hard” cases and that it always sits en banc. Id.

Not everyone is pleased that the Supreme Court is different from other federal courts. For an argument that the Supreme Court is increasingly unlike other federal courts in a fundamental way and that this is troubling, see Ashutosh Bhagwat, Separate but Equal?: The Supreme Court, the Lower Federal Courts, and the Nature of the “Judicial Power,” 80 B.U. L. Rev. 967, 968 (2000).

25. Thomas R. Kobin, Comment, The National Collegiate Athletic Association’s No Agent and No Draft Rules: The Realities of Collegiate Sports Are Forcing Change, 4 SETON HALL J. SPORT L. 483, 513 (1994) (“[I]t is the rare exception or fluke when an NFL player does not come out of a collegiate football program.”).

26. Mike Preston, It’s No-Win for Clarett, but NFL Vulnerable, BALT. SUN, Aug. 13, 2003, at 1C (“College football pretty much has the same rules and routine of the pro game . . . .”)
Yet the Choi and Gulati thesis is wrong as applied to football. It is not as far off base as it is for the judiciary, but it is wrong nonetheless. There is a highly respected award for the best college football player each year. It is called the Heisman Trophy.27 And yet “Heisman Trophy winners often fail in [the] NFL.”28 Indeed, of the sixty-eight winners in the history of the award, only eight have been inducted into the Professional Football Hall of Fame.29

Of course, not everyone agrees that the winner of the Heisman Trophy is, in fact, the best college football player in the land. There are inevitably extraneous factors, such as playing for a prominent school.30 Not everyone agrees with the winner of the Nobel Prize in Physics each year either. But keep in mind that only one person wins the Heisman Trophy each season. Whether or not the winner is always the best college player, it is clear that he is a very good college football player. Yet, “[i]nstead of dominating as professionals, most recent [Heisman] winners haven’t even been able to get into a game.”31 Of the ten Heisman Trophy winners between 1993 and 2002, seven, all of whom are offensive players, failed to gain a single yard in the 2003 NFL season.32

Once again, it is hard to be sure why great success at one level does not guarantee great success at the next. With the Heisman Trophy, the consensus seems to be that “[a] different type of player is successful at the college level as opposed to the pro level.”33

27. The Downtown Athletic Club of New York City has been awarding the Heisman Trophy, “college football’s most coveted and prestigious award,” since 1935. Ray Parrillo, All Leaders Are Worthy of Heisman, MACON TELEGRAPH, Dec. 10, 2002, at C3. The winner is chosen based on the votes of over 900 electors, including members of the media and former winners of the trophy. Id.
30. Mark Stewart, Free Fall: Heisman Winners Fizzle in NFL, MILWAUKEE J. SENTINEL, Dec. 8, 1999, at C1 (“Heisman Trophy winners usually have a few things in common: they play in a major conference, are members of highly successful teams, have name recognition at the beginning of the season and almost always play high-profile offensive positions—quarterback or running back.”).
32. Id. Charlie Ward, the 1993 winner, chose to play basketball over football; Danny Wuerffel and Rashaan Salaam, the 1994 and 1996 winners respectively, left the NFL after short careers; Chris Weinke (2000), Ron Dayne (1999), and Carson Palmer (2002) were on NFL team rosters in 2003 but failed to produce a single yard among them; and Eric Crouch, the 2001 winner, “quit football before his career started.” Id. Only Eddie George (1995), Charles Woodson (1997), and Ricky Williams (1998) had any real impact in 2003. Id.
33. Id. (quoting Gil Brandt, the former personnel director of the Dallas Cowboys and the NFL’s senior draft consultant).
ently in professional football, speed, size, and strength count for more than they did in college. 34 But whatever the reason, the evidence suggests that the effort by Choi and Gulati to find a Heisman Trophy winner among federal circuit court judges is unlikely to lead to uncovering great Supreme Court Justices.
WHO JUDGES THE JUDGES?

JOHN V. ORTH*

QUIS CUSTODIET IPSOS CUSTODES?1

In the routine performance of their judicial functions, judges are judged, in the loose sense of the word, by the litigants and by the general public, the latter usually informed through the reports of journalists. These two groups represent the ultimate consumers of the judicial product: the litigants directly as parties to the decision, the public indirectly as parties affected by the precedent established. On the judgment of these two groups over time rests the public reputation of the individual judge; on the cumulative judgment of all the judges rests the public reputation of the judicial branch of government.

In addition, judges are judged (still in the loose sense of the word) by the legal profession: lawyers who practice in their courts and other judges who rely on the precedents established by their decisions. Judges on appellate courts are also judged by law professors and students who study the written opinions that accompany judicial decisions.2 On these judgments rest the professional reputation of the judges.

In the case of a judicial decision appealed to a higher court, the deciding judge is judged, in a stricter sense, by appellate judges, who have the power to affirm or reverse the decision; occasionally the appellate judges comment on the work of the lower court, expressing praise or blame.3 Decisions of a final court of appeal cannot be formally reviewed by any other court. As Justice Robert H. Jackson observed of the judges of the U.S. Supreme Court, “We are not final because we are infallible, but we are infallible only because we are fi-

---

1 Juvenal, Satires VI, at 347.
2 The inclusion of a judicial opinion in law school teaching materials, such as casebooks, increases a judge’s reputation and salience. For a study of the factors that go into making a “leading case,” see Richard Danzig, Hadley v. Baxendale: A Study in the Industrialization of the Law, 4 J. LEGAL STUD. 249, 274-76 (1975). Trial court judges also produce written opinions, but it is rare for one to be included in legal educational materials.
3 It is not only trial judges whose decisions may be appealed; the decisions of appellate judges may be appealed to a still higher court. Judges on an intermediate appellate court also judge one another when a case is reheard by the court en banc. See, e.g., FED. R. APP. P. 35 (providing for rehearing en banc by order of a majority of the circuit judges in regular service whether on the suggestion of a party or not). The North Carolina Court of Appeals may be the only multimember appellate court in the United States without provision for rehearing en banc. See generally John V. Orth, Why the North Carolina Court of Appeals Should Have a Procedure for Sitting En Banc, 75 N.C. L. REV. 1981 (1997).
nal.”4 Of course, a later Supreme Court may criticize the judges’ reasoning in a prior case and overturn the precedent the case established, and judges, even on the highest court, may express contemporaneous judgments on their fellow judges in concurring or dissenting opinions.

Aside from these ordinary judgments on their day-to-day decisionmaking, judges are judged, in the strictest sense of the word, in cases of allegations of misconduct.5 Punishment for misbehavior ranges from relatively mild sanctions by a “judicial conduct commission” to impeachment and removal from office.6 Itself a legal proceeding, impeachment is governed by strict standards. The U.S. Constitution, for example, lists as impeachable offenses treason, bribery, and other high crimes and misdemeanors7 and provides for trial by the Senate, with removal from office by vote of two-thirds of the Senators present.8

Formal procedures for judicial impeachment and removal are the necessary corollary of the security of judicial tenure. Judges are not guaranteed employment, only protection against improper discharge. When judges served “at pleasure,”9 as in colonial America or in England before 1700, judges who misbehaved—however that was defined—were simply removed by the sovereign without the need to show cause. Once judicial tenure was made secure by the English Act of Settlement10 and American constitutions,11 judges served during “good behavior” and were removable only for good cause shown.

---

6. Punishment for official misconduct may include disqualification from holding further office. See, e.g., U.S. Const. art. I, § 3, cl. 7; N.C. Const. art. IV, § 4.
8. See id. art. I, § 3, cl. 6.
9. Durante bene placito, “during the good pleasure,” were the Latin words that originally described a common law judge’s term of office. David M. Walker, The Oxford Companion to Law 384 (1980). An office held by this tenure resembles a tenancy at will in the law of real property, terminable by either party at any time.
10. 12 & 13 Will. 3, c. 2, § 3 (1701) (“Judges’ commission[s] shall be made quamdiu se bene gesserint [so long as they shall behave themselves well], and their Salaries ascertained and established; but upon the Address of both Houses of Parliament it may be lawful to remove them.”). This Act applied to common law judges in England, but not to judges in the colonies. For the later history of judicial tenure in England, see David Lemmens, The Independence of the Judiciary in Eighteenth-Century England, in The Life of the Law: Proceedings of the Tenth British Legal History Conference 125 (Peter Birks ed., 1993). For the years after 1880, see Robert Stevens, The Independence of the Judiciary: The View from the Lord Chancellor’s Office (1993).
11. See, e.g., U.S. Const. art. III, § 1 (“The Judges, both of the supreme and inferior Courts, shall hold their Offices during good Behaviour . . . .”); Mass. Const. of 1780, pt. 2, ch. III, art. I (“All judicial officers, duly appointed, commissioned and sworn, shall hold
A further occasion for judging the judges arises when it becomes necessary to decide on retention if the judge serves for a term of years or promotion if the judge is not already on the highest court. Common law judges were traditionally appointed for indefinite terms, as federal judges and the judges of a few state courts still are. Terms of indefinite duration began when judges were removable at will; they became of greater significance when judicial tenure was made more secure. Departing from the common law and federal precedent, most American states now provide for fixed terms for judges, usually of long duration.

In many states, judges are chosen by popular election rather than appointed by the executive. Deciding whether to reelect a judge who has completed a term of office is a political decision, similar to deciding on candidates seeking their first judicial office. If party labels are allowed and judicial elections are contested, partisan considerations may preponderate, notwithstanding judicial qualifications that will
also be a factor. With sitting judges, it is past judicial performance that is judged; with first-time candidates, it is past performance as a member of the legal profession, assuming the selection of judges is limited to lawyers.  

When judges are chosen by direct election, a judge serving on a lower court may become a candidate for a position on a higher court; in contested elections, the opposing candidate may be a higher court judge seeking reelection. When judges are chosen by appointment rather than direct election, the appointing authority may have the opportunity to promote a judge from a lower to a higher court. The U.S. Constitution empowers the President to nominate judges to the Supreme Court and to appoint them "with the Advice and Consent of the Senate." Before the President would nominate a judge on a lower court for promotion to the Supreme Court, it is inevitable that some judgment would be made concerning the judge’s prior judicial

16. The U.S. Constitution prescribes minimum age and citizenship qualifications for Congressmen, Senators, and Presidents, U.S. Const. art. I, § 2, cl. 2 (Representatives); id. § 3, cl. 3 (Senators); id. art. II, § 1, cl. 4 (Presidents), but imposes no particular qualifications for federal judges. State constitutions at first did not require professional qualifications for judges, and in the early national period nonlawyers served on state supreme courts. See, e.g., John Phillip Reid, Controlling the Law: Legal Politics in Early National New Hampshire 22 (2004) (explaining that two of the three members of the New Hampshire Supreme Court in 1798 "had been trained for the ministry and had no education in law"). In many states a requirement of professional training was added later. See, e.g., N.C. Const. art. IV, § 22 (amended 1980) ("Only persons duly authorized to practice law in the courts of this State shall be eligible for election or appointment as a Justice of the Supreme Court, Judge of the Court of Appeals, Judge of the Superior Court, or Judge of District Court."); see also John V. Orth, The North Carolina State Constitution: A Reference Guide 118 (1993). Legal practice for a specific number of years may be a constitutional requirement for judicial office. E.g., N.J. Const. art. VI, § VI, ¶ 2 (amended 1978) ("The justices of the Supreme Court and the judges of the Superior Court shall each prior to his appointment have been admitted to the practice of law in this State for at least 10 years.").

17. Even when judges are chosen by direct election, vacancies between elections may be filled by appointment. See, e.g., N.C. Const. art. IV, § 19 (amended 1985) (authorizing appointment by the governor). As Professor Hurst cannily observed, "Especially where the term of office on the supreme court was a long one, vacancies by death or retirement from illness were frequent, for men were usually already of mature years when they came to the highest court." James Willard Hurst, The Growth of American Law: The Law Makers 134 (1950). In addition, a judge planning not to seek reelection might resign before the end of the term of office in order to allow the appointment of a successor, who could then run for election as an incumbent. In case of a vacancy on a higher court, it is of course possible that a judge serving on a lower court would be appointed.

performance.\textsuperscript{19} The Senate, too, would inevitably consider a nominee's prior service before confirmation.\textsuperscript{20}

Judging the judges is necessary for the integrity of the judicial branch, but it risks interfering with the proper discharge of the judicial function. Judges should decide cases according to the law and without fear of retribution. Over the long history of the common law, a balance has been sought between holding the judges accountable for proper performance and protecting them from improper interference. To protect them from harassment by private plaintiffs, judges have been immune from civil actions since at least the early seventeenth century.\textsuperscript{21} Increased security of judicial tenure—substituting service "during good behavior" for service "at pleasure"—was a response to perceived executive abuse. The English had insisted upon it, at least as to the judges at the center of the British Empire, in the constitutional settlement that followed the Glorious Revolution of 1688.\textsuperscript{22} The American colonists viewed the continuing appointment of their judges under the old form as a cause for separation. Among the grievances against King George III listed in the American Declaration of Independence was that "He has made Judges dependent on his Will alone, for the tenure of their offices, and the amount and payment of their salaries."\textsuperscript{23} The constitutions of the newly independent states and eventually of the United States remedied the omission.\textsuperscript{24}

Starting in the middle of the nineteenth century, states began to provide for direct election of judges.\textsuperscript{25} Direct election was part of the

\textsuperscript{19} During the presidency of Ronald Reagan, for example, the Justice Department is reported to have screened prospective judicial nominees. "If a candidate had previous judicial experience, that person's record would be carefully examined." Sheldon Goldman, \textit{Reagan's Judicial Legacy: Completing the Puzzle and Summing Up}, \textit{72 JUDICATURE} 318, 319 (1989).

\textsuperscript{20} For example, the nominations of John J. Parker in 1930 and Clement F. Haynsworth Jr. in 1969 were rejected by the Senate, at least in part, because of prior decisions in which they had participated as court of appeals judges. John Anthony Maltese, \textit{The Selling of Supreme Court Nominees} 56-59, 72-74 (1995).

\textsuperscript{21} See Floyd v. Barker, 12 Co. Rep. 23, 77 Eng. Rep. 1305, 1307 (Star Ch. 1607) (stating that without immunity, "those [judges] who are the most sincere, would not be free from continual calumniations").

\textsuperscript{22} See supra note 10.

\textsuperscript{23} \textit{The Declaration of Independence} para. 9 (U.S. 1776). The grievance may have been not so much that colonial judges were removable at the King's will as that they were not removable at all by the colonial assemblies. Hurst, \textit{supra} note 17, at 123.

\textsuperscript{24} See \textit{supra} notes 10-11.

\textsuperscript{25} Although Vermont led the way with direct election of some judges from 1777, the first state to provide for the direct election of all judges was Mississippi in 1832; New York followed in 1846. Hurst, \textit{supra} note 17, at 122. "Within ten years fifteen of the twenty-nine states which then made up the Union had followed New York. Every state which entered the Union after 1846 stipulated the popular election of all or most of its judges." Id. On the origin of the North Carolina provision for the direct election of judges, see John V.
Jacksonian impulse toward greater democracy, but it also reflected a recognition of the lawmaking role of common law judges. Although powers were separated in the new constitutions and legislative power was conferred on an elected legislature, judicial power—at least in a common law system—included a residual role in lawmaking. Common law judges not only resolved disputes, but they also established precedents to guide later decisions. True in England as in the United States, judicial lawmaking became more obvious and self-conscious in the circumstances of the New World. “The common law of England,” Justice Joseph Story once observed, “is not to be taken in all respects to be that of America.” American judges, state and federal, decided what parts of the common law to accept and in some cases changed the received common law.

Orth, Tuesday, February 11, 1868: The Day North Carolina Chose Direct Election of Judges, 70 N.C. L. Rev. 1825 (1992). Recently the trend has reversed:

A growing number of states have begun to back off from the pure elective principle. In the twentieth century, some states have adopted the so-called Missouri plan. Under this scheme the governor appoints judges, but his choice is restricted. A commission made up of lawyers and citizens draws up a list of names and gives it to the governor. The governor must choose from the list. The judge serves until the next election, then runs for reelection on his or her record. That is, the judge does not run against anybody; the public is simply asked to vote yes or no.

LAWRENCE M. FRIEDMAN, AMERICAN LAW: AN INTRODUCTION 85 (2d ed. 1998).

26. See Caleb Nelson, A Re-Evaluation of Scholarly Explanations for the Rise of the Elective Judiciary in Antebellum America, 37 AM. J. LEGAL HIST. 190, 207-10 (1993). Although Nelson emphasizes popular concern about the power of judges to interpret legislation and to refuse to enforce unconstitutional statutes, the lawmaking power of common law judges was also great; indeed, in antebellum America, state judges relied on the common law in reaching their decisions at least as much as on statutes and much more than on constitutions.

27. Van Ness v. Pacard, 27 U.S. (2 Pet.) 137, 144 (1829). Judge Thurman of the Ohio Supreme Court elaborated on this point:

The English common law, so far as it is reasonable in itself, suitable to the condition and business of our people, and consistent with the letter and spirit of our federal and state constitutions and statutes, has been and is followed by our courts, and may be said to constitute a part of the common law of Ohio. But wherever it has been found wanting in either of these requisites, our courts have not hesitated to modify it to suit our circumstances, or, if necessary, to wholly depart from it.


28. A state reception statute might provide broad guidance to the judges. For instance, North Carolina’s statute receives

so much of the common law as is not destructive of, or repugnant to, or inconsistent with, the freedom and independence of this State and the form of government therein established, and which has not been otherwise provided for in whole or in part, not abrogated, repealed, or become obsolete.

N.C. GEN. STAT. § 4-1 (2004). By contrast, the legislature was far more specific in determining which British statutes to accept. See sources cited infra note 31.
judges were lawmakers, judicial election was an obvious response to increase democratic accountability.\textsuperscript{29}

Individual common law rules could be altered by statute, but a more comprehensive response to judicial lawmaking, much discussed in the nineteenth century, was to eliminate judicial discretion insofar as possible by legislation codifying the law.\textsuperscript{30} This was, perhaps, more properly a judgment on the common law rather than on the common law judges. By means of a code, the legislature could decide for itself what parts of English common law to receive and when and whether to alter any of the received law—just as the legislature, rather than the courts, had decided which of the historic English statutes to recognize as so fundamental to the legal system as to be still in effect after independence.\textsuperscript{31} Unlike judicial election, codification met with only limited success in American states, due to professional hostility and legislative indifference.\textsuperscript{32}

\textsuperscript{29} Alan Watson has explained that

\begin{quote}
[\textit{f}or persons who believe in the sovereignty of the people, an independent judiciary presents a genuine problem: judges, once appointed with security of tenure may, even deliberately, frustrate the will of the people. One solution widespread in the United States is to have judges popularly elected and holding office only for a limited term.]
\end{quote}

\textsc{Alan Watson, The Making of the Civil Law} 155 (1981). Nations with a civil law system, such as France, attempt to solve the problem with a system of separate administrative tribunals, which makes the judges of those courts members of the executive branch of government. \textit{Id.}

In fact, direct election of judges did little to increase democratic accountability: “[P]opular election of judges became almost wholly a matter of form” due to the rise of political parties, which controlled the nomination process, and to the practice of executive appointment to fill vacancies. \textsc{Hurst, supra} note 17, at 128-34.

\textsuperscript{30} “By ‘code’ is meant here primarily a written work that is intended to set out authoritatively at least the principles and basic rules of a wide field of law, such as the whole of private law, commercial law, or criminal law, or of criminal or civil procedure.” \textsc{Watson, supra} note 29, at 100. “With codification, law becomes basically and primarily statute law . . . . The systematic and comprehensive nature of the code with supporting legislation makes law statute-oriented to an extent that is otherwise impossible . . . .” \textsc{Id.} at 168; see also \textsc{J.M. Kelly, A Short History of Western Legal Theory} 312 (1992) (describing the “leading idea” of the authors of the French Civil Code as an intent “to reduce as far as possible the interpretative and creative function of judges”). In the second half of the twentieth century, the contrast “between the civil-law judge, bound to the text of [the] code, and the common-law judge, free to construct new solutions for new cases,” was put in question. \textsc{Id.} at 407.


\textsuperscript{32} See \textsc{Shael Herman, The Fate and the Future of Codification in America, 40 Am. J. Legal Hist. 407, 414-18} (1996); \textsc{Charles M. Cook, The American Codification Movement: A Study of Antebellum Legal Reform} (1981), reviewed by \textsc{Robert W. Gordon, Book Review, 36 Vand. L. Rev. 431} (1983). Only Louisiana, influenced by its colonial experience under French and Spanish rule, adopted a comprehensive civil code. See \textsc{George Dargo, Jefferson’s Louisiana: Politics and the Clash of Legal Traditions}
Raising the stakes immeasurably in America was the emergence of the doctrine of judicial review. At least since the landmark decision in *Marbury v. Madison* in 1803, the U.S. Supreme Court has been the final arbiter of the constitutionality of the work of the other branches of government. Inevitably this has drawn the court away from the traditional judicial function of adjudication and into the political arena. Already in the 1830s, Alexis de Tocqueville recognized that “scarcely any political question arises in the United States that is not resolved, sooner or later, into a judicial question.” After the Judges Bill of 1925 made certiorari virtually the only route to the Supreme Court, the policymaking, as opposed to the more limited dispute-resolving, role of the court became increasingly obvious.

Once the judges had become removable only for cause, negative judgments on sitting judges—other than allegations of impeachable offenses—could not be formally expressed. In extreme cases, political dissatisfaction with particular decisions resulted in constitutional amendments. Or the legislature could attempt to preclude or reverse undesirable judicial decisions by adjusting the size of the court, reducing it as vacancies occurred or, more commonly, increasing the

---


33. *5 U.S. (1 Cranch) 137, 180 (1803).* Earlier instances of judicial review have been recognized in United States Supreme Court decisions. See *David P. Currie, The Constitution in the Supreme Court: The First Hundred Years, 1789-1888,* at 6-9, 20-23, 29-30, 37-41, 51-54 (1985).

34. 1 *ALEXIS DE TOCQUEVILLE, DEMOCRACY IN AMERICA* 280 (Phillips Bradley ed. & Francis Bowen trans., Alfred A. Knopf 1945) (1835). It is true that “political questions” are presented to the courts only after reformulation as “judicial questions” that are resolvable only by judicial means.


36. There had been an effort in the early years, fortunately unsuccessful, to make impeachment a means for the political branches to correct perceived errors by the judiciary. For example, Senator William B. Giles made the following remarks at the impeachment trial of Supreme Court Justice Samuel Chase:

A trial and removal of a judge upon impeachment need not imply any criminality or corruption in him. Congress had no power over the person, but only over the office. And a removal by impeachment was nothing more than a declaration by Congress to this effect: You hold dangerous opinions, and if you are suffered to carry them into effect you will work the destruction of the nation. *We want your offices,* for the purpose of giving them to men who will fill them better.

*Hurst,* supra note 17, at 136.

number of judges to allow for new appointments. In addition, a legislature could alter the jurisdiction of the court or regulate its time of operation.

The greater the difficulty of removing a judge, the greater the care exercised in judicial selection. On the part of the U.S. Senate, the consequence has been ever more careful scrutiny of nominees to the Supreme Court, which has led to the highest rejection rate for any appointive office requiring senatorial confirmation. Beginning in 1868, presidential nominations to all federal judgeships have been routinely referred to the Judiciary Committee for report to the full Senate. Public hearings began with the nomination of Louis D. Brandeis in 1916. Harlan Fiske Stone, nominated to the Supreme Court in 1925, was the first candidate-Justice to answer questions in person before the committee. Although undoubtedly fortuitous, the coincidence in 1925 of Stone’s confirmation hearings and passage of the Judges Bill giving the Supreme Court almost complete control over its docket is highly suggestive of the connection between increased Supreme Court power and heightened senatorial scrutiny of judicial nominees. Testifying by nominees has become routine since 1955.

Judges are judged by many different groups and for many different reasons. The public, generally, is concerned about the quality of justice. Lawyers are concerned on behalf of their clients; judges are concerned for the integrity of the judicial system. Academic commentators have made professional reputations a matter of serious study, while also occasionally engaging in the scholarly parlor game of listing the “greatest judges.” But the recent call by legal academ-

38. See generally John V. Orth, How Many Judges Does It Take to Make a Supreme Court?, 19 CONST. COMMENT. 681 (2002).
39. See Ex parte McCardle, 74 U.S. (7 Wall.) 506, 513-15 (1868) (holding that Congress has power pursuant to the Constitution to eliminate Supreme Court jurisdiction in certain cases).
40. To delay the Court’s decision in Marbury, Congress canceled the 1802 Term of the Supreme Court. See Act of Apr. 29, 1802, ch. 31, §§ 1-2, 2 Stat. 156.
41. MALTESE, supra note 20, at 2 (citing a 19.5% rejection rate).
42. THE OXFORD COMPANION TO THE SUPREME COURT OF THE UNITED STATES 770-71 (Kermit L. Hall et al. eds., 1992).
43. MALTESE, supra note 20, at 88.
44. See ALPHEUS THOMAS MASON, HARLAN FISKE STONE: PILLAR OF THE LAW 194-99 (1956). The questioning was actually at Stone’s request, to respond to critics who were concerned that his prior legal practice would make him too favorable to business interests. See id. at 194-95.
45. MALTESE, supra note 20, at 5, 109.
47. See, e.g., Bernard Schwartz, Supreme Court Superstars: The Ten Greatest Justices, 31 TULSA L.J. 93, 93-94 (1995) (beginning with the quotation: “The human animal differs from the lesser primates in his passion for lists of Ten Best.” (quoting DAVID
ics for empirical measures of judicial performance has not been motivated by merely scholarly interest in judicial reputations. Nor has it been designed to assist all those empowered to reelect or reappoint judges or to promote sitting judges to higher courts. Instead, it has been narrowly focused on the judgment of only one group, fellow judges, concerning judicial performance expressed in only one form, published judicial opinions, and for only one purpose, appointment to the U.S. Supreme Court.48

Assuming a norm of promoting to the Supreme Court a judge already serving on one of the federal courts of appeals, the new evaluators have focused their attention on the small cadre of judges thought to constitute the likely candidates.49 Use of the published opinions of these judges by other judges is treated as an indicator of positive judicial performance.50 The measures are intended as a means of guiding the President and the Senate in the exercise of their appointment power. Nominating and confirming a judge with a “low score” on the chosen measure would be open to criticism and perhaps rendered more difficult. What are wanted, in other words, are not so much empirical measures of judicial performance as empirical measures of political performance in the choice of Supreme Court Justices.

The search is supposedly for predictors of success as a Supreme Court Justice. The difficulty is not only that no other court is quite like the U.S. Supreme Court but also that success as a Supreme Court Justice is not easily quantifiable. Judges, or at least legally trained persons, may indeed be the best qualified to evaluate ordinary judicial performance. A form of legal connoisseurship is required that is not common knowledge.51 For this reason, common law

---

49. Such a norm would exclude other sources of recruitment that have in the past provided many notable Justices: state supreme courts (William J. Brennan), successful practice (Lewis F. Powell, Jr.), elective office (Hugo L. Black), government service (Robert H. Jackson), even law school teaching (Felix Frankfurter).
50. Judicial opinions are the characteristic work product of appellate judges. Other than federal court of appeals judges, the only other candidate class that could be considered under such a standard would be state appellate court judges.
51. Sir Edward Coke said as much hundreds of years ago when he lectured King James I: “[C]auses which concern the life, or inheritance, or goods, or fortunes of his subjects, are not to be decided by natural reason but by the artificial reason and judgment of
judges were routinely chosen from the legal profession. But what is true of ordinary dispute resolution and precedent-setting is not necessarily true of the policymaking demanded of a modern Supreme Court Justice. For that, other appellate judges are not necessarily the only, or indeed the best, judges. The judgment required may well be called political.

Proposals to institutionalize new measures for judging the judges must be situated in the context of the historic balance between protecting the judges from improper interference and holding them accountable for proper performance. Judging the judges has always been closely connected with judicial independence. The first formal process for removing judges from office, impeachment, was a necessary complement of “good behavior” tenure. Empirical measures of judicial performance relying on the judgment of other judges, if institutionalized, will necessarily affect the constitutional balance.

On the one hand, such measures will seemingly increase judicial independence by constraining the appointing powers in their decisions on judicial promotion. The judiciary would, de facto, become a party to the appointment process. While the politicians are unlikely to concede readily the implied loss of power, the general public, too, may well be loath to reduce political control of the judicial point of entry. The effectiveness of the judicial branch is largely dependent upon public acceptance of its legitimacy.

Furthermore, reliance on the judgment of judicial peers in judging the qualifications of potential Supreme Court Justices risks the stultification of the judiciary by promoting, in effect, consensus candidates. It would be ironic if a system that celebrates “great dissenters,” whose decisions later—sometimes years later—gain acceptance, would place a premium for promotion on the judgment of a majority
of contemporary judges. Metrics also smacks of formalism, which is reminiscent of the derided concept of “mechanical jurisprudence.”

On the other hand, empirical measures relying on the judgment of other judges will risk decreasing judicial independence by suggesting means for ambitious judges to increase the likelihood of promotion or for fellow judges to grant or withhold recognition in hopes of affecting the result. For more than half the history of the United States, the President refrained from promoting an Associate Justice to Chief Justice, in part, to avoid competition among the Justices for the President’s favor. At the time, it was thought that this would threaten to disrupt the normal functioning of the Court. In a more politicized environment, it could encourage ambitious judges to emphasize their agreement with the President’s policies—or disagreement, if the judge anticipates a change of leadership.

Emphasis on published judicial opinions is typical of law school lawyers; since the days of Dean C.C. Langdell at Harvard, American legal education has been centered on the judicial opinion. But however useful opinions are for pedagogic purposes, they are markers of judicial excellence only if they justify the right result. The decision rather than the opinion is the essential judicial function; what is decided is more important than what is said in explanation thereof. Subsequent judicial opinions may cite a prior opinion and may follow rather than reverse it; legal academics may include an opinion in casebooks and refer to it in scholarly articles; but if it explains, however elegantly, the wrong result, its author is not worthy of promotion—indeed, the opposite is true.

Reliance on the published opinion of the individual judge also threatens the collegial character of the judiciary. While not quite the invention of Chief Justice John Marshall, the unitary opinion of the

55. See generally Roscoe Pound, Mechanical Jurisprudence, 8 Colum. L. Rev. 605 (1908).

56. Edward Douglass White was the first Associate Justice of the U.S. Supreme Court to be “promoted” to Chief Justice, on the nomination of President William Howard Taft in 1910. Taft had forcefully rejected a suggestion that he nominate the senior Associate Justice, John Marshall Harlan: “I’ll do no such damned thing. I won’t make the position of chief justice a blue ribbon for the final years of any member of the court.” Walter F. Pratt Jr., The Supreme Court Under Edward Douglass White, 1910-1921, at 15 (1999) (internal quotation marks omitted).

57. C.C. Langdell was dean of the Harvard Law School from 1870-1895. He compiled the first casebook and established the case method of legal instruction. For a brief summary of Langdell’s “new legal world,” see Bernard Schwartz, Main Currents in American Legal Thought 346-53 (1993).

court owes its prevalence to him. From the first organization of the U.S. Supreme Court in 1790 until Marshall’s appointment as its head in 1801, the Justices conformed to English practice and delivered their opinions seriatim, one after the other. Under Marshall’s headship, the Court literally found a new voice, usually his own, and typically issued only one opinion. The opinion of the court was an institutional triumph, a “secret source” of judicial power. Emphasis on the personality of the author will lessen the weight of the judicial institution and likely encourage more frequent concurrences and dissents.

Judges have been judged since time immemorial. The Bible reports the case of the “unjust judge” who “feared not God, neither regarded man,” but who finally did justice to a poor widow because of her constant petitioning. Empirical measures that privilege the opinion over the decision it explains are not safe guides in the timeless search for a just judge. And privileging the judgment of the judges over that of the public at large and their elected leaders risks losing sight of the policymaking role of the modern Supreme Court.

61. Luke 18:2-6 (King James). The candid opinion accompanying this decision, whether or not it expresses the motivation of other judges, is unlikely to be widely cited: “Though I fear not God, nor regard man; yet because this widow troubleth me, I will avenge her, lest by her continual coming she weary me.” Id. 18:4-5.
In this Essay, I propose that judicial behavior is best understood as a function of the incentives and constraints that particular legal systems place on their judges. The approach is thus an economic one, but it is also commonsensical, has broad empirical support, and, of particular relevance to this Symposium, has strong implications for assessing judicial performance and performance-based criteria for judicial promotion. This Essay contains no original empirical research, but seeks to provide a framework for interpreting and guiding empirical studies of judicial behavior.

An immediate and important implication of the approach is that judicial behavior is likely to differ across national legal systems and indeed within a nation’s legal systems to the extent that components of the system (such as the different jurisdictions in the United States) differ in the incentives and constraints that they impose on judges. And still another implication is that the orthodox notion that judges merely interpret and apply law is unlikely to hold in all or even most legal systems. Another is that the criteria of judicial performance are relative to the incentives and constraints that determine judicial behavior. In some judicial systems, a judge’s reversal rate might be a critical performance criterion, while in others more weight would be placed on how often a judge’s opinions were cited by other courts or even on the political acumen exhibited by the judge in his opinions. It is a mistake to suppose that one performance criterion or set of such criteria should be applicable to all judges. I will pause from time to time in my analysis of judicial behavior to spell

* Judge, U.S. Court of Appeals for the Seventh Circuit, and senior lecturer, the University of Chicago Law School. This Essay was prepared for the Florida State University Law Review’s symposium, Empirical Measures of Judicial Performance. I thank Bryan Jenkins, Jonathon La Chappelle, Viktoria Lovei, Meghan Maloney, and Liss Palamkunnel for their very helpful research assistance and Andrei Shleifer for very helpful comments on a previous draft.
out in further detail the implications of my approach for performance evaluation.

My starting point in analyzing judicial behavior is the assumption that judges, like other people, are maximizers of their utility.1 That is, every judge has a utility function and tries to maximize the weighted utility of the arguments (representing preferences or desires) in the function. The utility function of the average person who is not a judge is likely to be dominated by income, leisure, family relationships, work satisfaction, and a concern for personal integrity, reputation, and felt achievement. The judge’s utility function is likely to be quite similar, but with somewhat different weights. Most people who seek or accept a judgeship probably derive more utility from leisure and public recognition relative to income than the average practicing lawyer does; the judge is also likely to be more risk-averse, since judicial incomes are lower but also more stable than those of practicing lawyers. Since no one is forced to be a judge, and the job is not to everyone’s liking by any means, there is self-selection—itself reflecting the play of incentives and constraints on human behavior—into the judiciary. And once selection has occurred, the incentives and constraints imposed by the structure and rules of the judicial career influence the judge’s behavior—for example, by inducing a greater pursuit of leisure or intellectual satisfaction relative to income—which in turn influences who is interested in becoming a judge.

The possible variations in incentives and constraints that could be brought to bear on judges are well-nigh infinite; to simplify the analysis, I shall analyze the behavior of judges in just a handful of possible configurations (incentive-constraint “packages”): private judges (that is, arbitrators); judges in career judiciaries such as one finds in most countries other than those whose legal systems derive ultimately from England; elected judges, such as one finds in most state courts in the United States; and U.S. federal trial judges, federal intermediate appellate judges (that is, federal circuit judges), and Supreme Court Justices. I will offer predictions based on the rational model of judicial behavior concerning the likely behavior of judges in these different systems and compare my predictions with actual, observed judicial behavior.

II. PRIVATE JUDGES (ARBITRATORS)

Arbitrators are selected by, or with the consent of, the litigants. An arbitrator who gets a reputation for favoring one side or the other

---

in a class of cases, such as cases of employment termination or disputes between investors and brokers or between management and unions, will be unacceptable to one of the parties in any such dispute, and so the demand for his services will wither. We can expect, therefore, a tendency for arbitrators to “split the difference” in their awards, that is, to try to give each side a partial victory (and therefore partial defeat). For this will make it difficult for the parties on either side of the class of suits in question to infer a pattern of favoritism. What is more, the pattern will be attractive to risk-averse disputants (because it will truncate both the upside and the downside risk of the dispute resolution process) and will therefore help to differentiate arbitration from adjudication. This is important because arbitrators, unlike courts, are not subsidized by the government; their fees and expenses must be defrayed by the disputants. The public subsidy of adjudication places them at a cost disadvantage vis-à-vis the courts. One way to overcome this disadvantage is to offer a distinctive service, and splitting-the-difference decisionmaking is such a service.

Arbitration offers something else attractive to risk-averse disputants: a lower error rate than juries, because arbitrators, when they are not lawyers, are business people who have experience relevant to the case at hand. This advantage is at least partially offset, however, by the fact that arbitration awards cannot be appealed (presumably to reduce the cost of arbitration and thus reduce the cost advantage of the courts), though they can be challenged in court on narrow grounds. Because of that offset, I am inclined to stress the splitting-the-difference character of arbitration in explaining the attractiveness of this substitute for adjudication as well as in elucidating the

2. “[C]ourts and juries are viewed as more likely to adhere to the law and less likely than arbitrators to ‘split the difference’ between the two sides, thereby lowering damages awards for plaintiffs.” Armendariz v. Found. Health Psychcare Servs., Inc., 6 P.3d 669, 693 (Cal. 2000); see also Bruce L. Benson, Arbitration, in 5 ENCYCLOPEDIA OF LAW AND ECONOMICS: ECONOMICS OF CRIME AND LITIGATION 159 (Boudewijn Bouckaert & Gerrit De Geest, eds., 2000); Christopher R. Drahozal, A Behavioral Analysis of Private Judging, 67 LAW & CONTEMP. PROBS. 105, 114-18 (2004); Estelle D. Franklin, Maneuvering Through the Labyrinth: The Employers’ Paradox in Responding to Hostile Environment Sexual Harassment—A Proposed Way Out, 67 FORDHAM L. REV. 1517, 1565 (1999) (finding that arbitrators alter “almost half” of the prior punishments in sexual harassment cases that come before them); Alan Scott Rau, Integrity in Private Judging, 52 S. TEX. L. REV. 485, 523 (1997); Donald Wittman, Lay Juries, Professional Arbitrators, and the Arbitrator Selection Hypothesis, 5 AM. L. & ECON. REV. 61, 81 (2003) (“[A]rbitrators tend to split the difference and consequently are much more likely [than civil juries] to find a verdict in favor of the plaintiff.”); Jane Spencer, Waiving Your Right to a Jury Trial, WALL ST. J., Aug. 17, 2004, at D1. But there is some support for the belief in findings that there is less variance in arbitrators’ awards than in jury awards. See Drahozal, supra, at 118. Since the lower end of the range of possible awards is truncated at zero, a reduction in variance is likely to reduce the average award. That would supply a motive for the contract party who was more likely to be sued than to sue for breach of contract to want an arbitration clause in the contract.
behavioral effects of privatizing judging, rather than to emphasize the more conventional differences between adjudication and arbitration.

Evaluating the performance of arbitrators is difficult, especially when, as is the common practice in commercial as distinct from labor arbitration, they do not write opinions. Although arbitration awards cannot be appealed, they can be challenged in court, on narrow grounds as I said; but presumably arbitrators whose awards are repeatedly vacated by the courts lose business, as judicial invalidation of the award creates added delay and expense for the parties, who, remember, bear the entire cost of arbitration. Also, lawyers observe the demeanor of the arbitrator in the hearings before him, and of course the outcome of the arbitration, and they form judgments, which they pass on to their clients, concerning the competence and biases (if any) of particular arbitrators.

III. CAREER JUDICIARIES: HEREIN OF THE ECONOMICS OF BUREAUCRACY

The career judiciaries found in countries whose legal systems do not have an English origin are, as the term “career judiciary” implies, systems manned by lawyers who make an entire career of being a judge. In contrast, most U.S. (and other Anglo-American) judges become judges only after a career in some other branch of the legal profession, such as private practice, prosecution, or teaching. In the U.S. federal judiciary, the average age of appointment to the district court has, since Harry Truman’s Presidency, varied from forty-nine to fifty-three, while that of circuit judges has varied from fifty to fifty-six. Obviously a lawyer first appointed to a judgeship in his forties or fifties is embarking on a second career.

A career judiciary is a part of the civil service. Appointment and promotion are by merit. Promotion is a critical feature of career judiciaries because a fresh law school graduate will naturally occupy the

---


4. Albert Yoon, Love’s Labor's Lost? Judicial Tenure Among Federal Court Judges: 1945–2000, 91 Cal. L. Rev. 1029, 1048 n.70 (2003). These are, I stress, averages. There is considerable variance on both sides. I was only forty-two when appointed (with no prior judicial experience) to the U.S. Court of Appeals for the Seventh Circuit; my colleague Frank Easterbrook was only thirty-seven when he was appointed; and some circuit judges have been appointed again with no prior judicial experience, in their sixties, though this is rare. Although some newly appointed or promoted federal judges have prior state or federal judicial experience (in particular, a substantial minority of federal court of appeals judges are promoted from the district court), most do not.
lowest rung of the judicial ladder when first appointed and will expect to rise to more responsible positions as he gains experience. I cannot see any important difference between a career judiciary and any other professional civil service, such as the diplomatic service or the armed forces. So the analysis of judicial behavior in a career judiciary should be essentially the same as the analysis of bureaucratic behavior in general, while bureaucratic behavior, in turn, should be similar though not identical to the behavior of employees in a large business firm; and let me start with that similarity.

The economic difference between an employee and an entrepreneur, in the sense of an independent business person, is that the employee does not sell his output; rather, he rents his labor to the employer. The employer tries to value each employee's output but recognizes that because the output of a firm is a team effort, only rough estimates are possible. And because of that roughness, the problem of agency costs arises: that is, the incentive of the employee or other agent to shirk if the cost to the employer of detecting some amount of shirking is greater than the benefit of preventing that shirking. The difficulty of valuing the employee's output leads employers to adopt proxies for that value, such as the employee's credentials and other input information, including the number of hours he works and the number of mistakes he makes. In general, the more costly it is to evaluate an employee's output, the more inclined the employer will be to substitute evaluation of inputs, such as credentials, hours, and care. These are costly as well as imperfect substitutes, and on both accounts one expects some shirking to remain uncorrected; stated differently, agency costs are unlikely to be eliminated entirely.

An obvious difference between a corporate and a government bureaucracy is that it is much more difficult to value the latter's output and therefore the value contributed by the individual bureaucrats. This increases agency costs, which in turn implies that the agents (the government bureaucrats) will have greater scope for pursuing their private ends than the employees of business firms. One private end is leisure; so one form that agency costs take in a bureaucracy as in a business firm is shirking. But in addition—and here we come upon a subtler difference between behavior in the two types of bureaucracy—the ideological character of the missions of many government agencies, in contrast to the strictly financial character of the profit-maximization goal of business firms, implies that agency costs in a government bureaucracy will take the form not only of shirking but also of “sabotage.” That is, employees will have a tendency to re-
define the agency’s mission to coincide with their personal ideological goals.\(^5\)

This is certainly a tendency of a judicial bureaucracy, given the political and ideological significance of adjudication. But there are proxies for minimizing these agency costs that resemble those employed by business firms. Credentials are one: grades in law school are a proxy, though like most proxies a rough one, for ability to perform well as a judge, which is one factor in the likelihood of actually performing well. Although the ultimate output of a judiciary—something that might be termed “legal justice”—is extremely difficult to value, the performance of a judge can be proxied by such things as backlog, reversal rate (consideration of which acts as a check on a judge’s minimizing his backlog by overly hasty decisionmaking—more on this later\(^6\)), judicial demeanor, complaints by litigants and lawyers, and, in a career judiciary, the quality of the judge’s rulings as evaluated by his superiors in the judicial bureaucracy.

One of the most important devices by which bureaucracies minimize agency costs is by laying down detailed rules for the bureaucrats to follow, since conformity to a rule is easier to determine than whether the bureaucrat is creative, innovative, imaginative, and so forth. Hence we expect and find that career judiciaries are found in legal systems that rely heavily on detailed codes rather than on the looser standards that are characteristic of common law systems. When a code sets forth a legal rule with great specificity, it is relatively easy to determine whether the judge is applying the code correctly; judicial agency costs are therefore minimized.

A career judiciary can be expected to be methodologically conservative and therefore unadventurous. Promotion in a career judiciary as in any other branch of the civil service depends ultimately on one’s ability to perform to the satisfaction of one’s superiors, and it is difficult to see how the supervisors in a career judiciary will benefit in their own careers from having bold, experimentally minded subordinates. It is not like a business firm, in which a division head’s hard-driving, innovative subordinates may produce increases in revenues and profits that will redound to his credit for having selected and encouraged those subordinates. Thus, we can expect the output of a career judiciary to display low variance, to be of uniformly professional

---


quality, but to be uncreative. It is no surprise that legal academics in nations that have career judiciaries are treated not merely as commentators on the law, as in Anglo-American legal systems, but as sources of law; the judges are unwilling to play such a role. With the structure of the judicial career and the heavy reliance placed on legal codes and treatise writers as sources of law, performance criteria that emphasized intellectual creativity and independence, political acumen, or even pragmatic insights into “law in action” would be misplaced. A judge who excelled in such dimensions would be stepping out of his designated role.

IV. PROMOTION PROSPECTS: CAREER VERSUS LATERAL-ENTRY JUDICIARIES

I have emphasized the role of promotion in constraining the behavior of judges in career judiciaries. In contrast, promotion is of limited significance in an Anglo-American-style lateral-entry judiciary. Most judges are not promoted at all, partly because judges tend to be appointed at a relatively advanced age, partly because there are very few rungs in the judicial ladder in most Anglo-American judiciaries, and partly because previous judicial experience is not required for appointment to even the highest rung. For example, although in the U.S. federal court system a significant fraction of intermediate appellate judges are appointed from the trial bench, most are not; and because there are many fewer appellate than trial judges, the great majority of district judges are not promoted. What is more, the salary and prestige differences between district and circuit judges are small, though the workload is lighter in the appellate court. And while at present almost all the Supreme Court Justices were federal circuit judges previously, there are so few Justices, and they serve for such a long time, that the percentage of federal court of appeals judges who

7. John Henry Merryman has noted that the civil law judge is a kind of expert clerk. He is presented with a fact situation to which a ready legislative response will be readily found in all except the extraordinary case. His function is merely to find the right legislative provision, couple it with the fact situation, and bless the solution that is more or less automatically produced from the union. The whole process of judicial decision is made to fit into the formal syllogism of scholastic logic. The major premise is in the statute, the facts of the case furnish the minor premise, and the conclusion inevitably follows.

JOHN HENRY MERRYMAN, THE CIVIL LAW TRADITION: AN INTRODUCTION TO THE LEGAL SYSTEMS OF WESTERN EUROPE AND LATIN AMERICA 36 (2d ed. 1985); see also Georgakopoulos, supra note 3, at 212 (noting that incentives faced by members of a career judiciary tend to discourage innovation); John Henry Merryman, The French Deviation, 44 AM. J. COMP. L. 109, 116 (1996) (discussing these effects as found in the French judiciary).

8. See Daniel Klerman, Nonpromotion and Judicial Independence, 72 S. CAL. L. REV. 455, 461 (1999) (“[T]he probability that a district court judge serving during the 1980s would be promoted to the court of appeals was only six percent.”).
becomes Supreme Court Justices is minuscule. Furthermore, while merit is not completely irrelevant to promotion in the federal court system (even promotion to the Supreme Court, 9 where political criteria dominate), it is not the dominant factor. In particular, the higher judges do not decide who among the lower judges shall be promoted. The result of all these factors is that promotion prospects cannot be expected to play a significant constraining effect on the behavior of American judges, in sharp contrast to the situation in a career judiciary. We must look elsewhere for the constraints.

V. ELECTED STATE JUDGES

The elected judiciaries of the U.S. states—and most of the states use some form of election to choose all or most of their judges—provide a striking contrast to the foreign career judiciaries. An elected judge is subject to constraints that have only attenuated counterparts in other types of judiciary. The first and most obvious is that, provided he is elected for only a limited term and therefore must stand for reelection, he is subject, as a tenured federal judge is not, to a form of performance review.

Second, and closely related, the elected judge has to be more sensitive to public opinion than a judge whose tenure does not depend on the whim of the electorate. Only a handful of cases, primarily those involving notorious crimes, will interest a significant portion of the electorate, but in those cases we can expect a systematic bias to creep in. For example, because only the most egregious murders are eligible for capital punishment, judges in a state that has capital punishment may tilt against capital defendants.10 In addition, we can expect elected judges to tilt more than appointed ones in favor of a litigant who is a resident of the judge’s state when the opposing party is a nonresident.11

What this means is that the evaluation of judicial performance in a system of elected judges is likely to be far different from that in a career judiciary. The electorate and one’s judicial superiors are very different types of performance evaluators with respect both to knowledge and to the criteria employed in the evaluation. Nevertheless,

9. See Lee Epstein et al., The Role of Qualifications in the Confirmation of Nominees to the U.S. Supreme Court, 32 FLA. ST. U. L. REV. 1145 (2005).
and notwithstanding all the criticisms that are made of judicial election, having to stand for reelection must have some effect in keeping judges on their toes, and there is no corresponding stick in the case of U.S. federal judges.

Third, the judge has to be able to raise money to conduct his electoral campaign, and the primary donors to judicial election campaigns are the lawyers who litigate in the judge’s court. If the lawyers on both sides in the principal practice areas—such as lawyers for medical malpractice patients and lawyers for medical malpractice defendants—gave equal amounts of money to judicial candidates, the situation would be much like that regarding arbitrators: the judge would have an incentive to steer a middle course in his rulings in such cases. But in fact the stakes in particular practice areas are often systematically asymmetrical, and in that event an elected judiciary is likely to display a systematic bias.

Putting these points together, we can see that elected judges are less independent politically than appointed ones, especially appointed judges with lifetime tenure. Yet this is not necessarily a bad thing, not only because of the spur to effort that not having tenure can impart, but also because the decisions of elected judges tend to be more predictable than those of appointed judges. This is consistent with, maybe even entailed by, the fact that elected judges are less independent; the independent judge is likely to have a more complex decision calculus, since he cannot just put his finger to the political wind. And as long as the populist element in adjudication does not swell to the point where unpopular though innocent people are convicted of crimes or other gross departures from the rule of law occur, conforming judicial policies to democratic preference can be regarded as a good thing in a society that prides itself on being the world’s leading democracy.

This point underscores the intimate relation between judicial behavior and judicial performance. If (and maybe it is a big if) one takes the existence of an elective judiciary to signify a legitimate democratic preference for aligning judicial and popular attitudes more closely than in a nonelective system, then a judge who defies public opinion is not only a judge unlikely to be reelected; he is, it can be argued, however paradoxically, a bad, even a usurpative, judge. The other side of this coin, however, is that the more uniform is public opinion, the more important judicial independence is in safeguarding

12. For additional evidence, see Brace & Hall, supra note 10. See also F. Andrew Hanssen, Is There a Politically Optimal Level of Judicial Independence?, 94 AM. ECON. REV. 712, 717 (2004) (providing references).
minority rights. Thus Andrew Hanssen finds that judicial independence is most likely to be valued where political competition is intense, because “[b]y establishing an independent court, politicians currently in office make it more difficult for successors to alter the policies passed today.”

Oddly, although an elective judiciary is more democratic than an appointive one in the Anglo-American setting, it is not more democratic than a career judiciary in legal systems that do not have an Anglo-American origin. When legislative codes are detailed and judges are formalists in the sense of enforcing the codes as written rather than using them as merely the starting point for the development of legal standards, the democratic legislature is calling the legal tune and the judges really are just executing decisions made by democratic process.

A further, and I think clearly adverse, effect of an elective judiciary is that it limits the field of selection. Most people are temperamentally unsuited for electoral politics and in any event are not good at it, though they may have just the suite of abilities required in an excellent judge. The number of people who have both political and judicial talent (and taste—a judicial career is likely to attract risk averters, and a political career risk takers) is probably very small, and there may even be a degree of incompatibility between the two kinds of talent. The list of failed politicians who went on to become fine judges is, I believe, longer than the list of successful politicians who became fine judges, though I have not been able to document this point. But assuming it is correct, we can expect that other things being equal, an elective judiciary will be less able than an appointive one—unless, to repeat a previous point, conformity to popular opinion is deemed a plus rather than a minus in a judge. Of course, other things may not be equal; in particular, lifetime tenure may, as I suggested earlier, have a debilitating effect on effort. The other side of this coin, however, is that lifetime tenure is a highly valuable asset, which increases the real income of federal judges relative to state judges and so contributes to making a federal judgeship more coveted, broadening the field of selection.

15. All that is clear—and surprising—is that in a large sample of federal judges, forty percent had held some kind of political office before becoming a judge, and twenty-seven percent of those (so roughly ten percent of the total) had held elected office. Judges’ Opinions on Procedural Issues: A Survey of State and Federal Trial Judges Who Spend at Least Half Their Time on General Civil Cases, 69 B.U. L. REV. 731, 755-56 tbl.9.1 (1989).
VI. FEDERAL DISTRICT JUDGES

Thus far in my analysis the question of judicial behavior has not presented many mysteries. The incentives and constraints facing judges in the systems that I have been discussing are generally rather similar to those playing on more conventional economic actors. The picture changes when we turn to the appointed, life-tenured trial judges of the U.S. federal judiciary. For the typical district judge, as I pointed out earlier, the appointment is terminal: the judge is unlikely to be promoted and therefore unlikely to be constrained in his rulings by desire for promotion. He is also unlikely to resign and go into the practice of law or into some other line of work, even after reaching retirement age; most federal judges prefer to take senior status then, which allows them to judge part-time yet receive their full salary. And only criminal or other egregious misconduct or incapacity will get the judge removed involuntarily. Moreover, not only is the judicial salary the same for all district judges—there are no bonuses for outstanding performance—but a judge’s ability to cash in on his judicial reputation by moonlighting as a teacher or lecturer is very limited, as there are strict, low caps on outside earned income, other than book royalties. It seems, then, that the federal judicial career has been carefully designed to insulate the judges from the normal incentives and constraints that determine the behavior of rational actors, except for the relative handful of judges who are ambitious for promotion to the court of appeals; and I shall ignore them. So it is as if the federal judicial career had been configured to perplex economists!

The mystery is at the practical rather than the theoretical level. Most decisions that a person makes have no greater impact on his utility than the decision of a judge has on the judge’s utility. A person faced with a choice between two nearly identical items on a menu, such as a choice between two flavors of ice cream, cannot base the choice on the effect on his income or job security, yet his choice will be rational: it will be the choice that generates a larger net increment in his utility. But it may be very difficult to figure out why that particular choice is the one that has that consequence. And so it is with judges. The decision of a federal district judge will not affect his income or job security, but it will affect his utility in some other way—the question is in what way. The difference between this and the ice cream case is that the effect of the judicial decision on the judge’s utility cannot be reduced to a single dimension, such as taste. Deciding a particular case in a particular way might increase the judge’s utility just by the satisfaction that doing a good job produces, which is what we would like. But it might also do so by advancing a political or ideological goal, economizing on the judge’s time and effort, invit-
ing commendation from people whom the judge admires, benefiting the local community, getting the judge’s name in the newspaper, pleasing a spouse or other family member or a friend, galling a lawyer whom the judge dislikes, expressing affection for or hostility toward one of the parties—and the list goes on and on.

Not only is there no effective mechanism for punishing a judge who yields to such temptations; that he has yielded can in most cases not even be detected. A judge’s broad discretion in managing the timing and scope of the litigation before him, in ruling on objections to evidence, and in resolving factual disputes enables him often to so influence the factual premises of the decision that it will appear to proceed ineluctably from the facts. The process by which some facts are highlighted and others ignored or given little weight need not reflect a conscious endeavor to make the judge look good and reduce the likelihood of reversal. It may simply be the consequence of uncertainty opening the way to bias. If an arresting officer says one thing and the person he arrested says something else, the judge’s decision as to which one to believe is likely to be influenced, sometimes decisively, by the judge’s background (was he a prosecutor before he became a judge? a defense lawyer?). Similarly, if one thinks back to the extraordinary variance in federal sentences that prevailed before the promulgation of the federal sentencing guidelines, one will find it difficult to resist the inference that the most important considerations in fixing a defendant’s sentence within the limits permitted by law had nothing to do with legal analysis but everything to do with the judge’s attitudes toward personal responsibility and toward the deterrent effect of criminal punishment.

An employer who cannot evaluate an employee’s output directly will, as I noted earlier, tend to base hiring, salary, promotion, and firing decisions on observable inputs instead, that is, on ex ante signals of quality. So one expects—and finds—that the more secure the tenure of the judges in a particular legal system, and hence the more difficult it is to control their behavior, the more careful will be the screening for the job by the appointing authorities; there will also be more competition for it, which will widen the field of selection and generate greater information for those authorities. Assuming that people have generally stable preferences and that behavior has a strong habitual element, the older a person is when he is appointed to a job, the more predictable his performance in it will be. A lawyer who has performed successfully for many years in practice, demonstrating qualities of sobriety, good judgment, integrity, and other attributes that are important in a judge, will probably continue to display those qualities when the carrots and sticks of a legal practice are withdrawn. It is like firing a gun: the position and rifling of the gun’s barrel impart direction to the bullet, but momentum takes over in
guiding the bullet once it leaves the barrel, though wind or other environmental disturbances may deflect it from its initial path. In much the same way, psychological momentum may cause a judge to behave consistently with his previous behavior as a practicing lawyer or prosecutor or professor even though he is freer in his judicial position from financial incentives and constraints than in his prior positions.

Moreover, the fact I have been stressing—that when gross incentives and constraints are removed, a space is created for ones normally of only minor significance to determine the individual’s behavior—has an upside as well as a downside so far as conforming judicial behavior to social norms is concerned. People care about their reputation apart from purely instrumental effects; that is why rank orderings and prizes have psychological effects distinct from any career effects of being singled out from one’s fellows. For example, federal district judges are sensitive to the quarterly statistics compiled by the Administrative Office of the U.S. Courts showing how many cases the judge has had under advisement for more than a specified length of time—so sensitive that judges will sometimes dismiss cases at the end of a reporting period, with leave to reinstate the case at the beginning of the next reporting period, in order to improve their statistics. Judges also do not like to be reversed, even though a reversal has no tangible effect on a judge’s career if he is unlikely to be promoted to the court of appeals in any event. Because judges are sensitive both to backlog and to reversal—neither allowing their backlog to grow to inordinate length merely to reduce the probability of reversal nor allowing their reversal rate to soar merely to eliminate their backlog by making precipitate rulings—they are constrained to exercise a kind of care that is analogous to that of judges in a career judiciary.

This is a neglected point, so let me elaborate on it a bit. District court judges have heavy dockets; a judge might well have 500 cases pending before him. Most of these will settle or be abandoned without judicial intervention; but enough will remain that require court action to induce the judge to attend to them, lest his backlog become completely unmanageable. And yet he cannot be completely summary in disposing of these cases because then his reversal rate will rise to an embarrassing level. So backlog pressure keeps him working hard, and reversal threat keeps him working carefully.


18. In addition, Higgins and Rubin found that reversal rate has no effect on a judge’s chances of promotion. Higgins & Rubin, supra note 1, at 135-36.
But the result is a band, not a point; within the band the judge has discretion—greater discretion than that enjoyed by judges in career judiciaries. Therefore, one expects that personal factors—such as political or ideological concerns personal to the judge rather than embodied in the law, the kind of intellectual laziness that consists of acting on intuition rather than on analysis and evidence, and the delights of tormenting the lawyers that appear before them—will play a larger role in federal district judges' decisions than they play in the decisions of their counterparts in career judiciaries. Tormenting the lawyers perhaps especially plays a larger role because it neither affects the judge's reversal rate nor increases his backlog; on the contrary, it will reduce his backlog by inducing more settlements. In short, judicial agency costs—the costs of controlling judicial behavior—are higher in a system in which judges have secure tenure and identical salaries than in one in which their careers depend on their ability to satisfy their superiors' expectations. When agency costs are higher, the agent has more discretion to pursue his own goals whether or not they coincide with his principal's goals.

It is important to distinguish, however, between judicial agency costs and political judging. The many studies which confirm—what everybody knows but orthodox legal thinkers are loath to acknowledge—that the political party of the appointing President is a good predictor of a judge's votes in a wide variety of cases shows nothing more than that there is a large open area in American law, that is, an area in which conventional legal materials do not dictate the outcome and the judge is forced to make a policy judgment, inevitably influenced by political or ideological preferences. The judge may still be a faithful agent of the President who appointed him or, to the extent that political preferences (not partisan preferences but preferences concerning public policy) are legitimate tools of adjudication, of "the law." The problem of agency costs arises only when the looseness of the principal's control over the judge enables the latter to make decisions driven by a preference that is too personal, partisan, or idiosyncratic, to be legitimate.

19. See, e.g., Steven Lubet, Bullying from the Bench, 5 GREEN BAG 2d. 11 (2001).
Similarly, the fact that judicial decisions are sometimes influenced by the race, religion, gender, or other personal characteristics of the judge need not be an effect of agency costs, but may merely reflect the fact that people from different backgrounds are likely to bring different priors to their resolution of factual issues and to have different policy preferences because of differences in life experiences.

VII. FEDERAL CIRCUIT JUDGES

The analysis of federal circuit judges is broadly similar to that of district judges but with four main differences. First, the dual constraints imposed by backlog pressure and reversal threat are attenuated. The caseloads of circuit judges are lighter than those of district judges, so the threat of an unmanageable caseload looms less ominously; moreover, once a case has been argued to the appeals panel, there will be no further activity in the case until it is decided, which means that the size of the backlog does not affect the workload of the appellate judge, as it does of the district judge. And so few court of appeals decisions are reviewed by the Supreme Court that the threat of reversal cannot operate as a significant constraint on circuit judges’ decisions—and for the additional reason that many reversals by the Supreme Court reflect ideological differences rather than error correction (and therefore explicit or implicit criticism); this is less true of reversals of district judges.

Second, because appellate judges sit in panels rather than by themselves, there is a premium placed on cooperative behavior; the downside is the risk of factions and (though I believe this is quite rare in the federal judiciary) of log rolling (that is, vote trading).

Third, appellate judges have a greater opportunity to influence the direction of the law, on the model most famously of Learned Hand, than trial judges do. One reason for this greater opportunity—apart from the obvious one that appellate adjudication focuses far more than at the trial level on general issues of law rather than on factual or procedural issues specific to the particular case—is that, as I have just noted, the Supreme Court reviews only a minute percentage (currently less than one percent) of court of appeals decisions. Entire fields of law are left mainly to the courts of appeals to shape. Many court of appeals judges are not ambitious enough to influence the direction in which the law will evolve or to acquire the kind of reputation that court of appeals judges like Learned Hand and Henry Friendly acquired; and because the risk of reversal is so much lower,

and the reward for creative legal thinking greater, at the court of appeals than at the district court, we can expect these judges to weight leisure more heavily, and to vote their personal preferences more often than district judges do. Like trial judges, and indeed more easily, appellate judges can conceal the role of personal preferences in their decisions by stating the facts selectively so that the outcome seems to follow inevitably or by taking liberties with precedents.

As for the minority of ambitious appellate judges (their ambition manifested for example in their often fierce competition to obtain the ablest law clerks\(^22\)), the principal constraint is stare decisis, that is, the practice of adhering to precedent. In a rational-actor analysis, the constraining effect of precedent comes not from the fact that stare decisis is a sound policy but from the fact that a judge’s influence is dependent to a significant degree on his decisions being treated as precedent by other judges. If he is cavalier about adhering to precedent in his own decisions, he undermines the practice of stare decisis in general and the likelihood that his own decisions will be followed by other judges in particular.

Another, though overlapping, minority of court of appeals judges should be mentioned: those who by reason of prominence, political connections, race or ethnicity, or other factors have a real shot at being appointed to the Supreme Court. With very rare exceptions, the probability that a given court of appeals judge, however well placed he seems in the competition to be appointed to the Supreme Court, will actually be appointed is low; but if the judge attaches enormous value to being a Supreme Court Justice, the expected utility of such an appointment (most simply, the utility of the appointment multiplied by its probability) may influence behavior. Thus, a study found that after Robert Bork’s nomination to the Supreme Court failed, in part because of his extrajudicial writings (the largest component of the “paper trail” that did him in), the publication rate of court of appeals judges, after adjustment for other factors, declined precipitately.\(^23\)

A fourth factor that differentiates circuit from district judges builds on the earlier suggestion (fundamental to this Essay) that when gross incentives and constraints on behavior are removed, smaller ones can be expected to have a decisive effect. It is that professional criticism of judicial decisions can be expected to place some limitations on the exercise of judicial discretion and more so at the appellate than at the trial level. The principal product of appellate

---

judges is the judicial opinion, and judicial opinions are self-contained (or at least self-contained-appearing) texts readily accessible to professional critique. However, present-day professional criticism of judicial opinions is so heavily discounted by most judges as to have little influence on their behavior, and thus it fails as an effective constraint on judicial discretion.

The reasons for this discounting are threefold. There is first a sense among judges that academics and practicing lawyers alike simply do not understand the conditions under which judges work and that much of their criticism of judicial performance is therefore captious, obtuse, and unconstructive. Second, rather than being disinterested, a great deal of the current professional criticism of judicial opinions reflects either client interests (in the case of criticism by practicing lawyers, even when expressed in books or articles) or, in the case of academic criticism, the politics of the professor. Third, and related to the second point, critique of judicial opinions emphasizes opinions of the Supreme Court to the virtual exclusion, or so it seems, of opinions of the lower courts, even though those opinions vastly outnumber Supreme Court opinions. Of course, treatises and law review articles dealing with areas of the law in which the Court is not active are perforce concerned with the work of the lower federal courts. But there is a great difference, so far as professional criticism as an influence on judicial behavior is concerned, between citing a judicial opinion for some proposition and analyzing the opinion in depth.

The relation of the third point to the second lies in the fact that to a great extent the Supreme Court—especially in its constitutional decisions—which are the particular focus of critique—is a political court, so that the critique of its work is also to a great extent unavoidably political, creating the impression that the nation’s law faculties have become increasingly politicized.

The fact that conventional professional criticism of judicial opinions is faltering badly as a constraint on the behavior of federal circuit judges makes the development of quantitative criteria of judicial performance, which have received emphasis in this Symposium, a welcome one. Such criteria are less likely than the conventional verbal standards applied to judicial performance to be dismissed as political and are also more economical because statistics can compact a vast amount of information, as discursive critique cannot. But five caveats are in order.

First, judicial performance criteria should not be uniform across courts and judges; such criteria as backlog and reversal rate should, for reasons indicated earlier, play a larger role in the evaluation of district judges than of circuit judges.
Second, as explained below, even first-rate performance criteria may not be useful for determining whether a judge should be promoted.

Third, as with any numerical ranking system there is a danger that the competitors will be able to “game” it.

Fourth, as illustrated by the performance criteria for circuit judges that have received the most attention, such as out-of-circuit citations or the more inclusive quantitative ranking scheme developed and applied by Choi and Gulati, the choice of criteria depends on assumptions about the incentives and constraints that operate on circuit judges. The criteria just mentioned measure influence and prominence (for example, one of Choi and Gulati’s performance criteria is the number of times a judge is mentioned by name, and in another study Gulati and Sanchez rank judges by the number of their opinions published in casebooks), implicitly treating judicial creativity as a desirable characteristic of circuit judges. Not everyone will agree; and it would be a particularly dubious assumption to apply to career judges and, to a lesser extent, to any trial judges and any elected judges, as well.

And fifth, as critics of the *U.S. News and World Report*’s ranking of colleges and law schools are well aware, numerical rankings are questionable when the rankings are multidimensional, so that the weighting of the different dimensions becomes critical to the rankings. Those weightings tend to be arbitrary. The problem is particularly serious with respect to the ranking of judges, because there is no agreement on what are the most important aspects of judicial performance. If you happen to think that lucidity is an extraordinarily important virtue of a judicial opinion, this will affect your weighting relative to someone who thinks that explaining carefully to the losing party why he lost, or discussing even minor issues, or stating the facts in comprehensive detail, is a more important virtue in a judicial opinion.

I should note that Choi and Gulati do not use reversal rate as a performance criterion for circuit judges. That may seem a surprising

---

24. For the most complete study, see William M. Landes et al., *Judicial Influence: A Citation Analysis of Federal Courts of Appeals Judges*, 27 J. LEGAL STUD. 271 (1998).


omission. However, reversal rate and creativity are likely to be positively correlated, since a judge who is creating precedents rather than just following them can be expected to be reversed more often than the unadventurous judge.\(^\text{28}\) (The analogy is to the fact that home-run hitters tend to strike out more often than singles hitters.) Moreover, the effect of reversals is captured automatically in one of the performance criteria that Choi and Gulati emphasize—frequency of citation to a judge—because a decision that is reversed is highly unlikely to be cited.

**VIII. THE SUPREME COURT**

When we turn to the U.S. Supreme Court, the picture changes once again. Reversal risk falls to zero, but there is still a constraining effect from stare decisis and the possibility of political retribution in the form of legislation (in the case of statutory decisions) or constitutional amendments (in the case of constitutional decisions) nullifying an unpopular decision.\(^\text{29}\) Political retribution can also take the form of low-level harassment by congressional budget committees,\(^\text{30}\) and can come from the prospect of the appointment of new Justices, when vacancies arise, who will be unsympathetic to the existing ones. Indeed, because of the high visibility of the Court’s decisions, the political constraints operating on the Justices are probably greater than those that operate at lower levels of the federal judiciary. But the combination of those constraints with the lack of guidance that conventional legal materials provide in truly novel cases, which bulk disproportionately large in the Court's docket, makes the Supreme Court importantly a political body—much more so than the lower federal courts—so that analysis of the behavior of the Justices should parallel that of the behavior of conventional political actors, as is frequently argued.\(^\text{32}\) Yet they cannot be evaluated entirely by that stan-


\(^\text{29}\) See generally William N. Eskridge, Jr., *Overriding Supreme Court Statutory Interpretation Decisions*, 101 YALE L.J. 331 (1991) (discussing and providing empirical evidence on Congress’s role in overriding Supreme Court statutory decisions and the extent to which these practices affect Supreme Court decisionmaking). For a more skeptical view of the importance and existence of these effects, see Frank B. Cross & Blake J. Nelson, *Strategic Institutional Effects on Supreme Court Decisionmaking*, 95 NW. U. L. REV. 1437, 1451-59 (2001).


standard, because a rational Justice will give weight to precedent. The political character of the Court demonstrates that the criticism of a Supreme Court Justice for having political smarts is not valid; and indeed it is a legitimate criticism of the current Court that the aggregate political experience of its members is slight by historical standards.

This analysis supports Steven Goldberg’s criticism of Choi and Gulati’s proposal that Supreme Court Justices be chosen by a “tournament” of federal court of appeals judges. Just as the best appellate judge in a Continental judiciary might very well not be the best choice for a U.S. court of appeals, so a court of appeals judge might very well not be the best choice for the U.S. Supreme Court. One must always beware the “Peter Principle”: the tendency to promote a person beyond the level of his competence as a reward for competent performance at the next lower level.

The best way to study the tournament proposal, begun by James Brudney in his essay for this Symposium, is to apply Choi and Gulati’s criteria to the federal court of appeals judges who have become Supreme Court Justices and see whether the criteria are predictive of the judges’ performance as Justices. (Their criteria can easily be extended to judges of other courts, to bring Holmes, Cardozo, O’Connor, and others who came to the Court from state judgeships into the picture. Of course this would require developing good performance measures for Supreme Court Justices.

IX. CONCLUSION

Difficult as the question of judicial behavior may appear to be from the standpoint of rational-actor analysis, a careful consideration of the incentives and constraints that operate on judges in different types of judicial systems and a careful exploration of analogies between judges and other economic actors, such as conventional bureaucrats and elected officials, may supply satisfactory answers, or at

33. For evidence, see Youngsik Lim, An Empirical Analysis of Supreme Court Justices’ Decision Making, 29 J. LEGAL STUD. 721 (2000).
34. Justice O’Connor is widely regarded as the most politically astute of the current Justices.
35. See Choi & Gulati, Tournament of Judges, supra note 25.
37. For other reservations concerning the adequacy of Choi and Gulati’s methodology for selecting Supreme Court Justices, see Daniel A. Farber, Supreme Court Selection and Measures of Past Judicial Performance, 32 FLA. ST. U. L. REV. 1175 (2005).
39. For example, in my book, Richard A. Posner, Cardozo: A Study in Reputation (1990), I used out-of-state citations, as well as other numerical criteria, in an effort to determine Cardozo’s standing among state judges. Id. at 74-91 & 85 tbl.5.
least establish a framework for further research. The analysis can, moreover, be extended to other judicial personnel, including law clerks and magistrate and bankruptcy judges, who operate under different incentives and constraints from those operating on the types of judges discussed in this Essay. And the better the behavior of judicial personnel is understood, the greater the feasibility of developing sound criteria of judicial performance.
PULLING FROM THE RANKS?: REMARKS ON THE PROPOSED USE OF AN OBJECTIVE JUDICIAL RANKING SYSTEM TO GUIDE THE SUPREME COURT APPOINTMENT PROCESS

HON. BRUCE M. SELYA*

I. INTRODUCTION

Although politics and ideology long have played a role in federal judicial appointment and elevation,¹ that role has swelled in recent years. Ongoing Senate confirmation battles over federal court nominees and rampant speculation about potential retirements from the Rehnquist Court have morphed into mainstream news. One critique of this ongoing fascination with the appointment process is that it is fundamentally out of focus. The contemporary debate centers on predicting how a putative Justice might (or might not) tip the balance on hotbed political issues rather than on merit or qualification for judicial service. Consequently, the debate says very little about how

* Judge, United States Court of Appeals for the First Circuit; Adjunct Professor of Law, Boston College Law School; Adjunct Professor of Law, Boston University School of Law. B.A., Harvard College, 1955; J.D., Harvard Law School, 1958.

¹. See Erwin Chemerinsky, Ideology and the Selection of Federal Judges, 36 U.C. DAVIS L. REV. 619, 620 (2003) (observing that despite the curiosity that “every generation has the sense that it is the first to uncover that ideology has a role in the judicial selection process,” in fact, “[e]very President in American history, to a greater or lesser extent, has chosen federal judges, in part, based on their ideology”).

1281
those dimensions might be measured. Traditionally, the American Bar Association has produced a qualitative measure of merit (which it provides to the President, the Attorney General, and the Senate Judiciary Committee) that evaluates the integrity, professional competence, and judicial temperament of each federal judicial nominee and designates each nominee as well qualified, qualified, or not qualified. This system, though valuable in some aspects, has limited utility because its “thumbs-up/thumbs-down” approach provides meager information about the relative merits of the nominees. Moreover, its legitimacy as a nonpartisan measure has come under attack by researchers who suggest that it could be a disguised political device.

Two academics, Professors Choi and Gulati, have reacted to the perceived hyperpoliticization of the federal judicial appointment and elevation processes by calling for the development of “objective” measures for evaluating judicial talent. To this end, Professors Choi and Gulati have put forth a particularly provocative proposal for a “Tournament of Judges” (the “Tournament”)—an ongoing empirical contest among federal appellate judges that would purport to tabulate objective measures of judicial performance (such as opinion publication rates, citation rates, and frequency of dissent) and would rank judges according to their overall scores. In the authors’ construct, the highest-ranked judge would be offered up as the heir apparent to the next Supreme Court vacancy. Choi and Gulati apparently believe that their Tournament will produce one of two desirable results: either politics will take a back seat to merit or, failing that, a politically motivated nomination will no longer be able to masquerade as merit-based. Even short of Supreme Court appointment, they posit, the Tournament would infuse the federal appellate bench with an “otherwise absent external incentive” for excellence in performance. A ranking system would hold judges accountable to high per-


5. See id. For a follow-up article in which the Tournament’s architects run the numbers to show how such a competition would operate, see Stephen J. Choi & G. Mitu Gulati, Choosing the Next Supreme Court Justice: An Empirical Ranking of Judge Performance, 78 S. CAL. L. REV. 23 (2004) [hereinafter Choi & Gulati, Empirical Ranking].

6. See Choi & Gulati, Tournament, supra note 4, at 301.

7. Id. at 304.
formance standards because the potential reward of elevation and the reputational benefits of superstar status would motivate them to excel.\textsuperscript{8} Conversely, the risks of decreased peer respect and public embarrassment would ensure good work even from those judges who have no realistic prospect of elevation or who are indifferent to judicial celebrity.

On the surface, these dual objectives—merit-based elevation and increased incentive to perform—may have a patina of plausibility. But there is more here than meets the eye. This Essay examines whether these objectives are worth pursuing in the abstract and then considers whether objective measures are likely to produce the results bruited by the sponsors of the Tournament.\textsuperscript{9}

II. OBJECTIVE MEASURES OF MERIT AND THE SUPREME COURT APPOINTMENT PROCESS

In the abstract, the task of developing objective measures of judicial performance seems straightforward. Achieving objective measures would assist in the selection of Supreme Court Justices both by providing standardized information about merit—valuable in itself for identifying the best candidates—and by gleaning the extent to which ideological alignment, diversity concerns, or other non-merit-based factors might drive a particular process of nomination and confirmation. It is, however, instructive to peek beneath the coverlet.

A. Do We Need Objective Measures of Merit to Identify the Most Qualified Candidates?

The primary function of objective measurement systems (such as the proposed Tournament) is to provide a standardized set of data in order to foster informed decisionmaking. The assumption underlying the perceived need for such data is that the informal, ad hoc methods currently employed to determine which candidates make the final cut are somehow deficient or that they lead to bad results. That assumption is somewhat puzzling. Although the existing process may be discursive and sometimes opaque, no one has made the case that it produces disagreeable outcomes (that is, that it results generally in the appointment or elevation of unqualified jurists). By any reasonable measure, the Article III judiciary comprises an array of talented men and women. This can only mean that there is an underlying merit-based quality control system at work. This system operates effectively, if somewhat obscurely, to ensure that, politics aside, success-

\textsuperscript{8} Id.

\textsuperscript{9} Though I focus on the Choi/Gulati proposal, I take that proposal to be representative of a broader bid to introduce empirical measures of judicial performance into the federal judicial system.
ful candidates have passed a certain threshold of merit. Thus, the real impetus behind objective measures cannot merely be a desire to ensure that candidates are qualified.

In fact, the absence of debate about the merits of nominees quite likely indicates that an informal but highly merit-conscious system of preliminary screening is already in place. To illustrate their point that decisionmakers eschew merit-based discussion, Choi and Gulati point to examples of political opponents talking past one another (such as the situation in which a proponent of a candidate asserts that the candidate is highly qualified and, instead of challenging that assertion head-on, the opponent responds with her ideological objections to the candidate’s elevation). 10

This claim that politics too often drowns out the merits fails to acknowledge that Presidents rarely submit nominees who would not fare well under an objective ranking system and that, as a result of that preliminary vetting, most nominees who come before the Senate easily meet or exceed any reasonable set of merit-based benchmarks. Were a candidate’s qualifications miserable or even comparatively underwhelming, one would expect the candidate’s foes to pounce on the fact of mediocrity rather than to engage in ideological polemics. In most cases, however, the fact of nomination is shorthand for the achievement of a behind-the-scenes consensus about objective qualification, such that public debate shifts almost immediately to more contentious issues.

Indeed, structural pressures all but guarantee that a President will nominate candidates who are highly qualified in terms of intelligence, experience, and skill. By vesting the nomination power solely in the hands of the President, 11 the Constitution concentrates accountability in a single individual. The caliber of the nominees will reflect directly upon the President. Alexander Hamilton wrote in The Federalist No. 76 that “[t]he sole and undivided responsibility of [the President] will naturally beget a livelier sense of duty and a more exact regard to reputation. [The President] will, on this account, feel himself under stronger obligations, and more interested to investigate with care the qualities requisite to the stations to be filled . . . .” 12 In contrasting presidential appointment with the alternative of appointment by a multimember assembly, Hamilton observed that in the latter case “the intrinsic merit of the candidate will be too often out of sight.” 13

10. See Choi & Gulati, Empirical Ranking, supra note 5, at 26 & n.3, 27; Choi & Gulati, Tournament, supra note 4, at 321 & n.55.
13. Id. at 493.
Nor is there a mandate or even an implicit norm that requires a President to offer seats on the Supreme Court as a reward for being the best in terms of intelligence, skill, or service on a lower court. Therefore, respect for the nominating power and regard for the many nuanced components that enter into the selection process counsel against lodging merit-based objections unless a particular nominee is in fact poorly qualified. Merit operates only as a threshold, not as the ultimate determinant. One can argue for increasing the height of the threshold limit, but even if that is done, there still will remain a group of aspirants who can exceed it. It is nothing short of utopian to think there will be only one.

In sum, for all that can be said about the lack of mainstream discussion of merit, there exists no plausible basis for a substantiated claim that the present process fails to yield top-notch Justices. If an objective measurement system is redundant because it identifies roughly the same pool of candidates that historically have been considered and chosen, there will be very little tolerance for any costs that such an innovation imposes. The question, then, is whether the game is worth the candle.

B. Do Objective Measures of Merit Serve the Goal of Political Transparency?

The idea of developing objective measures of merit to direct Supreme Court appointment grows out of a desire to reduce the politicization that has increasingly plagued the process of nominating and confirming appeals court judges (and that threatens to embroil future Supreme Court nominees). This is a matter of preference, not of constitutional mandate. The Constitution assigns the responsibility of choosing Supreme Court Justices to the political branches of the federal government, and therefore it is unsurprising that ideology factors into the exercise of the nomination and confirmation powers. As a normative matter, that is as it should be. Even assuming that the extirpation of politics from the nomination and confirmation processes is a goal supportable on other grounds—and that is a stretch—I have no faith that any system of rankings would achieve it. From my perspective, political transparency—and not some false promise of liberation from politics—is the touchstone in considering the value of objective measures.

Seen in this light, the argument for the use of objective measures is that while we may not be able to eliminate the politics, we can at

---

15. See Chemerinsky, supra note 1, at 628 (arguing that “ideology should be considered because the judicial selection process is the key majoritarian check on an anti-majoritarian institution”).
least unmask the political subtexts that underlie conflicting claims that a particular aspirant is either an intellectual giant or an intellectual pygmy of the kind whom we consider to be qualified/not qualified to serve on the Supreme Court. The example that Choi and Gulati offer is a presidential nominee who ranks forty-second out of 160 active circuit court judges on the merit-based scale.\textsuperscript{16} In that situation, they assert, a neutral observer can conclude that some factor other than objective merit (say, ideology) is driving the selection.\textsuperscript{17}

That is hardly rocket science. Yet another equally likely scenario demonstrates how a ranking system may serve to provide an impervious cover for an essentially ideological choice. While a President may have some explaining to do if he nominates number forty-two, he will escape the burden of revealing his calculus if he picks a candidate from the top tier (say, from the fifteen top-ranked circuit judges), regardless of whether he bases that selection on the candidate’s ideological alignment. In that way, a President will be able to hide behind the very metric that the Tournament’s advocates have intended as a means of making his motivations transparent. Although Choi and Gulati submit that the “introduction of a norm to apply objective criteria will force politicians to provide more justification for their selection,”\textsuperscript{18} it may very well accomplish exactly the opposite result.

The likely proliferation of multiple “objective” metrics also may undermine the goal of political transparency. Merit, like beauty, often lies in the eye of the beholder. If decisionmakers do come to rely on a merit-based ranking system, the lure of potential influence on the appointment process doubtless will spawn a market of competing objective indices, each claiming greater accuracy. Savants, docents, and other interested parties will lose little time in developing formulas designed to reach particular results. Consequently, any ranking system will face constant criticism that it is a proxy for either political affiliation or ideological leanings rather than for merit. I fear that the jousting between warring indices would repoliticize the selection process at a level even further removed from the relative qualifications of judicial aspirants. The goal of political transparency will not be served by a band of cleverly designed ranking systems that merely masquerade as objective measures.

\textsuperscript{16} Choi & Gulati, \textit{Empirical Ranking}, supra note 5, at 28.
\textsuperscript{17} Id.
\textsuperscript{18} Id. at 30.
C. Are Objective Measures of Judicial Performance Apt Indicators of Qualification to Serve on the Court?

Even if one assumes that, in the abstract, an objective ranking system could assist in the selection of candidates for the Supreme Court and could improve the political transparency of the process, the fact remains that a ranking system is only as effective as its component criteria. Choi and Gulati posit that judicial merit can be measured by a composite of “productivity,” “quality,” and “independence.” To measure productivity, they factor in the number of majority and dissenting opinions a judge has written and the number of cases in which the judge has participated. To gauge quality, they seize on what they believe to be an already accessible market test for that attribute: looking at how often and how prominently a judge’s opinions are cited by other courts and academics. Finally, to gauge judicial independence, they reason that disagreement, particularly opposition to the opinion of a colleague appointed by a President of the same party, indicates that the judge has a mind of her own, and thus dissent rates constitute a valid proxy for this attribute.

While there is arguably some relationship between these three indices and judicial merit, I doubt that the correlation is close enough to justify placing much weight on a composite ranking, let alone to justify using such a ranking as the primary filter for narrowing the field of candidates for elevation. My concerns fall into two general categories. First, I sense a series of disjunctures between the stated criteria and the concept of judicial merit. Second, even if the proxies were acceptable, they would be inherently manipulable by the contestants. I consider each index in turn.

1. The Productivity Index

In evaluating the productivity index, my immediate concern is that this measure is underinclusive because it ignores a range of appellate activities. Though writing opinions is the primary work of the appellate judge, that activity captures only part of the occupation.

19. Id. at 42 (defining the terms productivity, quality, and independence for the purposes of their article).
20. For a more detailed description of this methodology, see id. at 42-47.
21. Id. at 48.
22. See id. at 61-67 (explaining the methodology for calculating the independence measure).
23. Professors Choi and Gulati readily acknowledge that the Tournament’s measures are flawed in many ways but dismiss these defects by downplaying what is in many ways the Tournament’s most attractive selling point—that it actually will tell us something about merit so that we can evaluate candidates on that basis. To do this, they inch away from the question whether the Tournament gets it right when it comes to identifying merit and instead insist that so long as the Tournament unmask political forces, it is a valuable tool. See id. at 35-36.
the interest of expediency, judges decide many cases by issuing judgments or memorandum orders, without full-dress opinions. Those devices are utilitarian, and their appropriate use should not be discouraged. Judges also regularly serve on duty panels that decide motions and other procedural matters—while it is not glamorous work, it is, nonetheless, necessary. At the request of the President or the Chief Justice, some judges undertake service on bodies such as the Sentencing Commission, the Judicial Panel on Multi-District Litigation, the Foreign Intelligence Surveillance Court, and committees of the Judicial Conference. This service often limits the number of cases that a judge hears in his own court. Finally, many judges are forced to restrict their sitting schedules in order to carry out the administrative burdens that are essential to the operation of the federal courts.

While the underinclusivity of the index is arguably correctable to some extent, it highlights the ease with which one can rig a ranking system to produce particular results by selectively including certain factors and excluding others. Moreover, this deficiency reminds us that ranking systems are scarcely masters of nuance; a particular candidate will fare well only if she fits the mold that the index constructs. Designing that index is a normative task that demands some agreement about what merit means. This enhances the likelihood that competing versions of merit—versions that are likely inseparable in certain ways from judicial philosophy (read “ideology”)—will produce competing indices, thereby clouding an observer’s ability to discern merit.

My next criticism is that the productivity factor is deceptively straightforward. By placing productivity on a pedestal, the ranking indulges an assumption that we want Supreme Court Justices who are docket-movers. That is a wholly untested notion of merit. Indeed, the index assumes that those who write more slowly or less often are inefficient, lazy, distracted, or just plain incompetent. That assumption is a canard: it unfairly demeans judges who subscribe to a philosophy of restraint or those who believe that publication for publication’s sake tends to confuse the law.24 The point is that the index has its biases—and more troubling still, those biases may not be immediately apparent.

Manipulability is also a concern. Judges, to varying degrees, exercise control over their own numbers. If a judge wanted to increase her productivity score, she might work harder, take more cases, and produce more opinions. So long as this were done without sacrificing quality, this option would bring about a higher ranking while helping

the administration of justice. But a judge might avail herself of other options to effect a score increase without accomplishing any systemic gains. She could, for example, sacrifice quality for quantity, dash off gratuitous concurrences or dissents, shy away from demanding cases in search of easy prey, or use a host of other tactics. These myriad opportunities for “strategic gaming” of the numbers would further debilitate the effectiveness of the index as a proxy for merit.

Last—but far from least—there are real costs to introducing new behavioral incentives into an institution in which justice has long been thought the cardinal goal. Hasty and unnecessary opinions generally make for bad law, and manipulative behavior may erode both the credibility of the system and the collegiality that is so necessary to the effective functioning of an appellate court. The point is that even though a productivity index may incentivize genuine improvements, there is an equal or greater risk that it will encourage behaviors inimical to the development of decisional law and to the legitimacy of the federal courts.

2. The Quality Index

The assertion that consumption indicates quality may have some truth in it because judges and academics concerned with producing coherent rationales are unlikely to rely on poorly reasoned ones. At least two distortions, however, will likely interfere with the quality signal.

The first problem with culling a quality score from citation rates is that all citations are not created equal. A citation may signal that the authoring judge produced an apt statement of a commonly used standard or a path-breaking approach to a complex area of the law. Absent a sophisticated coding system—one that would doubtless involve subjective judgments—the rankings would not record such nuances.

A second problem (and perhaps one that we should associate more generally with all supposedly objective measures of judicial performance) arises from the fact that subsequent authors will come to rely upon the rankings themselves, particularly the quality index, as a shorthand for quality. This will create a sort of self-fulfilling prophecy. Judges who rank high on the quality index will likely earn respect in the legal community, thereby increasing the probability that other judges and academics will cite their opinions. Future citation rates will then say less about the quality of a cited opinion and more about the entrenched reputation of the original writing judge. This illustrates the Achilles’ heel of the Tournament proposal. Rankings tend to take on a life of their own, and many people tend to rely on them without knowing much about what they really signify. Put
bluntly, rankings too often disengage “the thing” from “the thing signified” and thereby frustrate the objective (informed, substantive deliberation) for which they purport to supply a foundation.

Moreover, the quality index is ripe for manipulation. Any judge worth his salt will tell you that there are ways to write opinions that make citation more likely. Judges can boost citation rates by writing longer opinions, publishing opinions that would otherwise go unpublished, eschewing quotations, taking controversial positions, or reaching for novel issues at the margin of a case. Accordingly, frequency of citation sometimes may signal better strategy rather than better quality.

3. The Independence Index

The independence index aims to capture the intellectual independence of the judge by measuring his willingness to disagree with his colleagues. This is pure fiction: rates of dissent are an invalid proxy for judicial independence. Many judges write fewer dissents because they ascribe a relatively high value to the institutional good of courts speaking with a single voice, and they are willing to work toward developing a template to which the entire panel can subscribe. A low dissent rate for such a judge is a badge of honor, not of shame. In all events, it tells us nothing about his independence.

Even if a judge’s rate of dissent provides some tiny amount of information about her independence, it is not clear that independence in that sense is a desirable trait. After all, there plainly is a point at which dissenters cross the line from enriching thought into either intellectual preening or obstructionist polemicism. The sponsors of this proxy thus overlook the obvious danger in encouraging dissent for its own sake (or more precisely, for the sake of a better independence score).

Furthermore, Choi and Gulati have conceded that a particular ranking system could elect to place a negative weight on dissent. This concession highlights another of the many points at which subjective choices—choices of judicial philosophy and ideology—make their entry into a supposedly objective measurement system. And, finally, the likelihood of manipulability is very high. Incentivized manipulation always troubles me—but it troubles me particularly in this context because it encourages dissent for the sake of dissent and, in the bargain, threatens both collegiality and the clarity of the law. The cruel irony of the independence index is that the act of measurement threatens to destroy the thing measured.

26. Id. at 62.
I might add that this effect is far worse than the perverse incentive that Choi and Gulati attribute to the current appointment system. They asseverate that judges now use voting, opinion writing, and dissent to signal their willingness, if elevated, to pander to the ideology of the appointing administration. In my experience, this seldom occurs.

4. Recapitulation

The bottom line is that I cannot credit the claim that these three indices constitute objective measures that aptly gauge generally accepted notions of judicial merit. Moreover, the costs of the Tournament proposal are yet to be calculated—and those costs must be justified against whatever meager benefits an actual Tournament might provide. I now turn to those costs.

D. The Siphoning Effect of Employing Objective Measures of Merit as an Initial Screen

The federal appellate judiciary is not a densely populated institution. Having narrowed their field to include only active federal appellate judges, Choi and Gulati place approximately 160 candidates in their Tournament. Although they frankly acknowledge that there are normative and historical objections to choosing such an exclusive pool as the starting point for Supreme Court appointment—a after all, two of the nine sitting Justices come from other venues, and Justice Souter was a sitting federal appellate judge for only a day—they attempt a test run of the Tournament by further narrowing the sample to ninety-eight judges (eliminating latecomers and those who did not remain active through June of 2003). Even more problematic than this apparent elitism is the fact that a tournament system aggrandizes attributes on the basis of ease of measurability rather than relevance to what really makes an ideal Supreme Court Justice. Other dimensions, such as temperament, integrity, and worldliness, are left by the wayside. Unless one actu-

27. See id. at 34.
28. See id. The authors concede that the threat of nomination blocking provides a check on overly explicit signaling; they contend, however, that judges engage in “stealth signaling” to avoid proclaiming ideology so strongly that opinions and dissents arouse the other side to “muster its resources to block [the candidate].” Id.
29. Choi and Gulati have conceded that an actual Tournament system would not provide a perfect or even a nearly perfect measure of judicial merit. See, e.g., id. at 35-36; Choi & Gulati, Tournament, supra note 4, at 312.
30. See Choi & Gulati, Empirical Ranking, supra note 5, at 28.
31. See id. at 40; Choi & Gulati, Tournament, supra note 4, at 318-19.
32. Choi & Gulati, Empirical Ranking, supra note 5, at 40-41.
33. See id. at 35-36.
34. See id.
ally values publication, citation, and dissent rates as the most exalted of all attributes—and I doubt that any of us who are not braindead would commit to that proposition—one must acknowledge that a tournament system may skip over candidates who, considered holistically, are the most desirable.

In sum, the Tournament proposal aims to circumscribe the field of candidates and even to limit the bounds of discourse based on a bobtailed version of merit. Self-imposed restraints of this order require heightened justification, and it does not seem to me that the elucidation of the role of politics in the selection process is sufficiently important to warrant an artificial scheme that uses a three-attribute measure of merit to tip the selection scales in favor of particular candidates. I do not place such great faith in these objective indices, nor have I discovered in the Tournament proposal any sustainable normative argument that rates of publication, citation, and dissent are the most important attributes of a Supreme Court Justice.

III. THE EFFECTS OF OBJECTIVE MEASURES OF MERIT ON THE PERFORMANCE OF THE FEDERAL APPELLATE JUDICIARY

Having explored the implications of employing objective measures of merit in the Supreme Court selection process, I now turn to a secondary consequence of a selection system that relies heavily on such measures: the incentive effect, that is, the influence that the ongoing operation of a ranking system that holds itself out to be merit-based will have on the everyday performance of the federal appellate judiciary. Choi and Gulati posit that promotional and reputational motivations will compel judges to modify their behavior in response to a ranking system. They find this result to be desirable because they believe that the Tournament identifies excellence, and therefore the incentive to score high in the Tournament will extract optimal performance from federal judges. I cannot quarrel with their assumption that judges will react to publication of these statistics. I am less comfortable, however, with the premise that these incentives will produce a better, fairer, and more efficient judicial system. In a democracy that relies heavily on a system of checks and balances, more accountability is generally welcomed. But the Tournament imposes accountability for the wrong actions—like a failed health care system that compensates physicians for cloning things but not for taking care of patients. It also threatens the judicial independence that, under our system of government, is constitutionally guaranteed. Thus,

35. See Choi & Gulati, Tournament, supra note 4, at 300, 313-14.
36. See id. at 304.
the Tournament’s avowed mission to remedy “the lack of incentive to seek promotion” invites scrutiny.

I begin with a brief review of the mechanism by which Choi and Gulati anticipate this transformation will occur. The basic idea is that federal appellate judges—on the whole, a group of high achievers (and, thus, competitive)—will conform their behavior to the criteria used in the Tournament in pursuit of promotion (or at least the bragging rights that accompany a favorable ranking). The public, in turn, will benefit from the amount of attention and care the judges devote to the measured tasks. Even if we set aside all questions pertaining to (1) the competitiveness of judges and (2) the relationship between the proposed criteria and the development of admirable judicial traits, there remains the question whether the promised result justifies the introduction of an external pressure on judicial decisionmaking.

Choi and Gulati exert little effort in addressing the implications of an incentive system for judicial independence. They write:

Whatever other objections exist, the one that we do not see room for is the argument that the tournament would hurt judicial independence. If anything, the pressures that appellate judges may currently feel to attract political sponsors by making decisions that please those sponsors would be eliminated. Indeed, if there is an objection to our system at all, it is that judges will be made too independent under it. The tournament will thus have eliminated one of the few popular checks on an otherwise independent judiciary.

This is more a distraction than an answer to the question. Among other things, it disparages the emphasis that the Founders placed on the independence of the judiciary. The Federalist No. 78 heralds the “independent spirit in the judges which must be essential to the faithful performance of so arduous a duty.” This spirit is embodied in Article III’s tripartite guarantee of independence, which comprises lifetime appointment, undiminished compensation, and exclusive vesting of the judicial power of the United States.

Despite their importunings to the contrary, Choi and Gulati’s Tournament places this independence in the crosshairs. By their own measure, the Tournament system is successful if it exerts a force on federal judges that causes them to modify their behavior in a way

37. Id. at 300.
38. See id. at 314-15.
39. Id. at 320-21.
40. THE FEDERALIST NO. 78, at 504-05 (Alexander Hamilton) (The Modern Library ed. 1937) (“The complete independence of the courts of justice is peculiarly essential in a limited Constitution.”).
41. Id. at 508.
42. U.S. CONST. art. III.
that brings it more in line with the ranked criteria. They boast that this will provide a degree of accountability.

In this context, a certain measure of accountability is healthy. We already require federal judges to comply with ethical rules, subject them to impeachment for high crimes and misdemeanors, and allow Congress to control the federal courts’ jurisdiction, budget, and rule-making authority. But Choi and Gulati’s proposed method for ensuring stellar judicial performance is a horse of a different hue. It not only conflicts with the Founders’ apparent distaste for perpetual job evaluation,43 but it also has its roots in a desire for personal advancement. Under the Tournament model, the impulse that spurs judges to perform to the best of their abilities comes not from a desire to administer perfect justice but, rather, from a desire to pursue promotion or indulge in an ego trip. This is not the kind of accountability that we should aim to inculcate.

One response to the admonition that objective measurements will jeopardize judicial independence is of the “lesser evil” brand—servitude to the rankings will relieve judges of their existing servitude to politics, so that they no longer will feel the need to modify their behavior to suit the ideologies of a sitting President. The hope is that the primary incentive will be to score better on a merit index, so that the system will reward good performance rather than cynical politicking. The trouble with this argument is that it presumes—without an iota of proof—that the new incentive will eliminate the old one. What is far more likely, however, is that a judge attentive to her Supreme Court ambitions will be tempted to serve two gods: ideological purity and the Tournament rankings.

The second response to the judicial independence concern is that regardless of what motives a Tournament system arouses, one cannot call its incentives pernicious so long as the result—better judicial performance—is beneficial. According to this thesis, the end justifies the means: We should not care if judges serve themselves or serve justice; the only thing that matters is that they produce timely opinions, get cited, and register regular dissents.

With respect, that is smoke and mirrors. Motivation is material because it is connected to the quality of the judicial product and, consequently, to the strength of the combined indices as a proxy for merit. As pointed out earlier, incentive-based judicial products may be of lower quality and, because the Tournament system is not designed to detect lower quality products and treat them differentially, these incentive-based products quite probably will further erode the proxy value of the three criteria. Over time, judicial rankings will say

43. See id.
less about actual merit and more about agility—the ability to game
the system.

The inevitable secondary effect of the Tournament system thus
undermines its primary purpose. The initial impetus for the meas-
urement system is the search for judges who exhibit certain qualities
that comport with the ideal of a jurist who will serve justice well, but
because this is a reward-offering search, a tournament may instead
lead us to judges who, like Pavlov’s dogs, respond well to incentives
and who understand what manipulations will create an apparent
match to that ideal.

IV. THE INSTITUTIONAL EFFECTS OF OBJECTIVE MEASURES OF MERIT
ON THE FEDERAL APPELLATE JUDICIARY

If the President and the Senate were to embrace an empirical
ranking system to assist them in their respective nomination and
confirmation responsibilities, other consequences would follow. I fo-
cus here on the implications for relations within and between circuits
and the potential impact on the legitimacy of the federal appellate
judiciary.

A. The Effect of Competitive Ranking Systems on Intracircuit and
Intercircuit Relations

It does not take a sophisticated analysis to point out that intro-
ducing an assessment system that pushes judges to vie for position
against one another will create potentially unhealthy competitive
pressures. Within circuits, the parade of horribles would look some-
thing like this: division of labor would become a constant bone of con-
tention, particularly with respect to the assignment of opinions; the
independence index would spawn superfluous dissents, thereby caus-
ing damage both to collegiality and to the rule of law; and the rank-
ings would create an explicit pecking order for an already competi-
tive process of hiring law clerks. In short, this sort of competition
would have unhealthy consequences for the way that we work with
one another.

A second set of consequences would imperil intercircuit relations
and could well lead to the isolation of the law of each circuit. Notably,
the particular quality index that Choi and Gulati have selected to
plug into their composite formula primarily measures “outside cita-
tions,” meaning citations by judges from other courts.44 Because the
Tournament ascribes a high value to cross-circuit citation but no

44. Choi & Gulati, Empirical Ranking, supra note 5, at 71 (stating that the quality
index consists of an adjusted figure representing the number of outside citations for a
judge’s top twenty opinions and the number of invocations).
value to intracircuit citation, a judge can cite to an intracircuit colleague without increasing her score but may be reluctant to cite an intercircuit colleague, who is also her competitor. After all, doing so would give the cited judge a boost in the rankings. This consequence may discourage citation to opinions from other circuits that, while not controlling authority, are illuminating and persuasive, even though such reliance would contribute to uniform development of federal law. In the worst-case scenario, this undesirable strain on the conversation between circuits as they encounter novel questions of law may even encourage artificial circuit splits or uncertainties about whether circuits agree or disagree on particular points of law.

Another consequence arises from the eventuality that the rankings will generate notions of prestige that attach to the reputation of the circuit at large. Circuits that do not host a high-ranking Tournament competitor may command less respect or attention from other judges, lawyers, academics, and clerkship aspirants. Those circuits may then suffer from circumscribed influence on the development of the law. Though effects of this kind are subtle, they are as difficult to control as they are to detect.

The more general point is that rankings tend to assert a peculiar power in our society, whereby they come to define our entire perception of the things they measure instead of providing a limited set of information about those things. And, in practical terms, it is impossible to cabin the influence of rankings to the narrow purpose for which they were designed.

B. Life at the Bottom and Judicial Legitimacy

Even though ranking systems are inherently relative, they inevitably give rise to labels that quickly become cemented to the name and reputation of each ranked member. If decisionmakers credit judicial rankings by relying upon them, the public and the legal community may come to perceive a judge ranked at or near the bottom not just as relatively less productive, less respected, or less independent than his peers but also as a rotten judge. This presumption of incompetence would emerge regardless of the fact that virtually the entire membership of the federal appellate bench is first-rate. In a ranked list, someone has got to wind up at the bottom.

The rankings also could warp the legitimacy of bottom-tier judges. Low rankings and their attendant consequences may alienate the judges who receive them. A system that effectively creates a second-class judiciary could come to define entire careers. I fear that an entrenched ranking system would make it far too easy to forget that every federal appellate judge has been entrusted by the President and the Senate to hold the federal judicial power for life.
V. CONCLUDING REMARKS

One thing can be said with certainty about Choi and Gulati’s Tournament of Judges: it catapults us into an engaging metaphysical experiment. In the end, however, I think that it sends us down an unsightly path. The now-popular spectacle of university hierarchs scrambling to game the *U.S. News and World Report* rankings would caution us to avoid that path.

In my view, the judiciary would do well to keep its pikestaffs at parade rest and eschew the jousting that Choi and Galuti invite. Objective measures such as they describe are likely to supply information that is only marginally beneficial in a system that performs well in identifying talented candidates. That information is as likely to obscure political motives as it is to expose them. Moreover, the suggested measures may be wildly inexact proxies for merit. The gains that a Tournament would provide are modest at most—and the cost of them, should they materialize, is the infliction of substantial harm.

To cinch matters, artificially grafting the proposed measures of judicial performance onto the current selection process misconceives the concept of Supreme Court elevation by portraying a Court vacancy as the ultimate reward to which objectively deserving federal appellate judges are entitled. That is as wrong as wrong can be. We ought to understand nomination and confirmation as a complex search for an individual who will best serve the nation at a particular point in time. Suitability for service on the Court cannot be reduced to a matter of baseball card statistics. From that perspective, the effort to line up the federal judiciary into tidy rows, ordered exclusively by a handful of objectively measurable considerations, interferes with what is a much broader process. What is more, the business of the federal appellate judiciary is the administration of justice through the exercise of Article III jurisdiction. As it stands, a Supreme Court vacancy is an infrequent interruption of that business. Thus, it makes little sense to advertise the unlikely prospect that political lightning will strike as a legitimate daily preoccupation for federal appellate judges. It makes even less sense to hope that federal judges will indulge in such a distraction.

In the last analysis, federal appellate opinions are not applications for employment on the Supreme Court. Scoring judges as if that were the endgame takes too narrow a view of a President’s prerogatives while at the same time encouraging the membership of an independent judiciary to subordinate judicial wisdom to the whims of personal ambition. Objectively speaking, the Tournament proposal strikes out on its pitch that it will produce merit-based decisionmaking and transparency.
DO JUDGES BEHAVE AS *HOMO ECONOMICUS*, AND IF SO, CAN WE MEASURE THEIR PERFORMANCE?
AN ANTIPODEAN PERSPECTIVE ON A TOURNAMENT OF JUDGES

RUSSELL SMYTH*

| I. INTRODUCTION | 1299 |
| II. WHAT IS *HOMO ECONOMICUS*? | 1301 |
| III. WHAT MOTIVATES JUDGES TO PERFORM? | 1302 |
| A. Financial Incentives | 1302 |
| B. Promotion Prospects | 1304 |
| C. Respect of Colleagues | 1306 |
| D. Influence | 1307 |
| IV. HOW SHOULD WE MEASURE AND RANK JUDICIAL PERFORMANCE? | 1309 |
| A. Opinions Published | 1309 |
| B. Citations | 1313 |
| C. Reversals | 1315 |
| D. Judicial Gowns or Academic Robes? | 1316 |
| V. IS IT APPROPRIATE TO MEASURE AND RANK JUDICIAL PERFORMANCE? | 1317 |
| VI. INCENTIVE EFFECTS OF MEASURING AND RANKING JUDICIAL PERFORMANCE | 1325 |
| VII. CONCLUSION | 1329 |

I. INTRODUCTION

The purpose of this Essay is to examine Choi and Gulati’s concept of designing a “tournament of judges,”1 with particular focus on whether it would be useful in the Antipodes (Australia and New Zealand). Focusing on Australia and New Zealand offers an interesting comparison with the United States and allows one to flesh out some of the issues associated with designing a tournament in countries with seriatim opinion writing and no judicial conference. There are some similarities in terms of promotion arrangements (for example, in New Zealand most members of the Court of Appeal are promoted from the High Court2) and some differences (for example, in Australia promotion to the High Court comes from one of several courts or from private practice3). There are also differences in the appointment

---

* Professor, Department of Economics, Monash University. BEd (Hons); LLB (Hons); MEd, Monash University; Ph.D, London University. I thank Mitu Gulati, Ingrid Nielsen, and Richard Posner for helpful comments on an earlier version of this Essay. The discussion of the psychological dimensions of measuring judicial performance was developed with substantial assistance from Ingrid Nielsen.


2. Note that in New Zealand, the New Zealand High Court is for all intents and purposes the second highest court, akin to the courts of appeals in the United States, while the New Zealand Court of Appeal is the final court of appeal, akin to the United States Supreme Court.

3. In Australia, the High Court of Australia is the final court of appeal. Below the High Court of Australia is the Federal Court of Australia, which is broadly similar to the United States courts of appeals and the state supreme courts. If a judge is to be promoted
process. One of Choi and Gulati’s main rationales for formulating a tournament of judges is that the process of appointment to the United States Supreme Court is so politicized. This is less so in Australia and New Zealand. The judges of the High Court of Australia and the New Zealand Court of Appeal are appointed by the Governor General, who is the representative of the Queen of the United Kingdom, on the advice of the relevant national government. Thus it is effectively the government that makes the decision on who to appoint, but without the series of Senate hearings into the background of the candidate, which characterizes the appointment process in the United States.

Having said this, the appointment process in Australia and New Zealand still lacks transparency. Governments in both countries are under no obligation to consult with state governments (relevant in the Australian case), the judiciary, or the profession more generally, and with the exception of a few spasmodic anecdotal accounts, exactly how appointments are made is largely shrouded in mystery. The basic notion that some of that nontransparency could be cleared up if the government would provide hard data about judicial performance to back its claim that a particular judge is the best candidate for a position still holds.

There is ongoing debate in Australia and New Zealand on the issue of how judges should be selected, with some arguing for public scrutiny of the opinions and attitudes of judges similar to what occurs in the United States. There is also continuing controversy over whether the High Court of Australia should be more representative of the population (for example, in terms of gender, there has only ever been one female judge on the High Court of Australia) and the implications this would have for judicial performance. Former Chief Justice of the High Court of Australia, Sir Anthony Mason, has stated that while a more representative judiciary “may assist in

to the High Court from a lower court, it will typically be from the Federal Court or one of the state supreme courts.


5. See, e.g., Palmer, supra note 4, at 40-52; Winterton, supra note 4, at 185-88.


maintaining public confidence in the administration of justice, . . . it is essential that [it] be achieved without any diminution in the quality of judicial performance. The insistent demand for enhanced judicial performance requires the appointment of those who are best qualified." 8 If Australia was to go down that track, the notion of a tournament of judges would become more pressing. In Australia, there is also continuous discussion of the merits (or lack thereof) of introducing performance standards raised by law societies, 9 to which quantitative indicators of the sort suggested by Choi and Gulati would be a useful contribution. One difference, however, between what Australia would use the tournament for and the manner in which Choi and Gulati set up their tournament is that the debate in Australia tends to center on prescribing minimum standards to get the below-average judges up to scratch rather than setting indicators designed to identify the highfliers.

II. WHAT IS HOMO ECONOMICUS?

The issue whether judges behave as homo economicus (economic man) is a useful starting point, because the premise underpinning Choi and Gulati’s tournament of judges is that judges respond to incentives. 10 The view was long held that judges were immune to self-interest. 11 From the late 1990s, however, sparked mainly by Posner’s seminal work on judicial decisionmaking, 12 a few articles started to appear that argued that judges, like the rest of us, are self-interested and respond to incentives. 13 While the notion that judges are self-interested welfare maximizers has not been unanimously accepted, 14 it has taken root in law and economics’ treatment of judicial behav-

10. Choi & Gulati, supra note 1, at 305.
ior. The notion of self-interest has always been at the center of neoclassical economics. Bowles and Gintis note that generations of students have been told that “[t]he strength of the neoclassical paradigm . . . lies in its hardheaded grounding in a general model of self-interested action.”\textsuperscript{15} The role of self-interest in the economics literature, though, has evolved over time. Bowles and Gintis further suggest that “self-interested behavior underlying neoclassical theory is artificially truncated: it depicts a charmingly Victorian but utopian world in which conflicts abound but a handshake is a handshake.”\textsuperscript{16} The reality, of course, is more complex. In the 1970s, a new \textit{homo economicus} started to emerge in the principal-agent, the economics of information, and the radical political economy literature that reflected these complexities in human behavior. “The new economic man is not a Victorian gentleman: he is uncompromisingly thorough in pursuing objectives, and often he is less benign.”\textsuperscript{17}

Williamson suggests that the new \textit{homo economicus} will engage in a “full set of \textit{ex ante} and \textit{ex post} efforts to lie, cheat, steal, mislead, disguise, obfuscate, feign, distort, and confuse.”\textsuperscript{18} If \textit{homo economicus} behaves like this, one’s initial reaction might be that it is ridiculous to suggest that judges behave in this fashion. To take an extreme example, most would agree that it is absurd to postulate that judges (or most judges) are likely to expend effort on stealing. But, at the same time, this is also true for most members of the broader community. Judges, along with members of the broader community, might, however, engage in shirking of work effort, which represents a form of theft of time. And judges might engage in behavior designed to mislead or disguise their real output. The point is that \textit{homo economicus} will behave in an opportunistic manner to achieve his self-interested objectives. It is important to take this into account when designing incentives to improve performance and mechanisms to monitor their implementation.

III. WHAT MOTIVATES JUDGES TO PERFORM?

A. Financial Incentives

Before we can say anything about how to measure judicial performance and how such measurements will affect judicial behavior, it is important to say something about what motivates judges. If it is accepted that judges do indeed respond to incentives, an examination

\textsuperscript{16} \textit{Id}.
\textsuperscript{17} \textit{Id}. at 84.
\textsuperscript{18} OLIVER E. WILLIAMSON, THE ECONOMIC INSTITUTIONS OF CAPITALISM 51 n.8 (1985).
of the motives of judges will shed light on the most appropriate incentive structure. In Australia and New Zealand, as in the United States, associate judges of the same court get paid the same amount. Given there are often sizeable differences in output across judges, wages do not reflect marginal product. The compensation of judges does not depend on their rulings, and the quality of the precedents set by the judge does not result in any direct pecuniary benefit.\textsuperscript{19} From time to time, judges have expressed dissatisfaction with their compensation in the face of mounting workloads in both Australia and the United States.\textsuperscript{20} In Australia, the issue of judicial compensation was a major issue at the last election in 2004, with the Federal opposition stating that if elected, it intended to abolish the present noncontributory judicial pension scheme and replace it with community-standard superannuation arrangements (the opposition was defeated). In part, this was a response to the blowout in the unfunded liability for judicial pensions. The law societies responded by arguing that this would have an adverse effect on the ability to attract the best people to the bench.\textsuperscript{21} There is some evidence that pension entitlements influence the retirement decisions of judges. Squire\textsuperscript{22} and Zorn and Van Winkle\textsuperscript{23} find that pension eligibility has a statistically significant positive effect on the propensity to retire from the United States Supreme Court. Spriggs and Wahlbeck reach the same conclusion for the United States courts of appeals,\textsuperscript{24} although Barrow and Zuk\textsuperscript{25} and Hall\textsuperscript{26} find that financial incentives are less important in the lower federal courts and state supreme courts, respectively. Maitra and Smyth find that pension eligibility has a significant positive effect on the retirement decision in the High Court of Australia.\textsuperscript{27}


\textsuperscript{22} Peverill Squire, \textit{Politics and Personal Factors in Retirement from the United States Supreme Court}, 10 POL. BEHAV. 180, 185-86 (1988).


\textsuperscript{26} Melinda Gann Hall, \textit{Voluntary Retirements from State Supreme Courts: Assessing Democratic Pressures to Relinquish the Bench}, 63 J. POL. 1112, 1130 (2001).

Nevertheless, it is highly questionable whether there is much scope to improve judicial performance through improved financial rewards. The reason is that it is unlikely that many judges are motivated by financial returns. Monetary compensation only represents a small portion of the returns to judging, and it is a diminishing proportion the higher the court, because on higher courts the nonmonetary returns are higher. While there are no studies on this point, the decision to take a judicial appointment is reflected in the facts. The opportunity cost of judging, purely in terms of monetary income, is high. Most judges would have been earning considerably more in private practice and, as such, give up substantial monetary incomes to join the bench. Similarly, most judges could earn more by resigning from the bench and returning to private practice or even acting as a consultant to a major law firm, because of the premium that the prestige of being a former judge would attract. While inadequate pension entitlements are sometimes given as a reason for judges retiring early and returning to private practice in Australia, on the High Court of Australia at least, the examples of judges having to linger in office because of inadequate funds or having to return to private practice to replenish their coffers all occurred long ago. Anecdotal evidence suggests that now, by the time they accept judicial appointment, most senior barristers in Australia have substantial sums invested in private pension plans and have accrued large numbers of assets, so they are not reliant on their judicial pensions to survive in retirement.

B. Promotion Prospects

Judges rarely, if ever, would admit that desire for promotion was a motivating factor. However, as Schauer notes in the U.S. context, while judges may be circumspect in admitting it, “it is hardly implausible to suspect that many trial judges desire to become appellate judges, and that most judges of intermediate appellate courts (including the United States Courts of Appeals) desire to become

30. Macey, supra note 13, at 630 n.9.
32. In the early years of the High Court of Australia, there is anecdotal evidence of judges remaining in office in the face of failing health for financial reasons. O'Conner (1903-1912) was unable to retire despite suffering from chronic illness from 1907 onward because he was pensionless; he eventually died in office. Griffith (1903-1919) suffered a stroke in 1917 and sat on few cases in his last two years on the Court, but he refused to retire because he had insufficient funds; he eventually retired when the government passed legislation granting him a pension.
33. Schauer, supra note 13, at 623.
judges of courts of last resort.”

34. There is some evidence that judges in Japan and the United States are motivated by prospects of promotion. Cohen presents empirical evidence that desire for promotion influences the behavior of district judges in the United States. 35 Taha shows that judges with higher prospects of promotion are more likely to publish their opinions. 36 In a series of articles, Ramseyer and Rasmusen find that promotion prospects influence the behavior of lower court judges in Japan. 37

Desire to be promoted is also likely to be a motivating factor in New Zealand and the United Kingdom. Both nations have career-based judiciaries, with most judges of the English House of Lords being promoted from the English Court of Appeal and most judges of the New Zealand Court of Appeal coming from the New Zealand High Court. It is interesting to note that, in the case of both New Zealand and the United Kingdom, existing studies suggest that governments take into account both political factors (independence exercised by the judge) and indicators of the quality of judicial decision-making (such as the proportion of cases in which the judge has been reversed) in deciding whether and when the judge should be promoted. 38

In Australia, promotion is less likely to be a motivating factor because the prospects of promotion are fewer, at least to the High Court of Australia. There have been only forty-four members of the High Court, including the current justices; and they have been appointed from a variety of places, including lower judicial office, politics, and private practice. Less than half (twenty-one) have held prior judicial office. Of those who have held prior judicial office, most (seventeen) have been appointed from state supreme courts, primarily from either Victoria or New South Wales, with the others being pro-

34. Id. at 631-32.
promoted from the Federal Court. Because there is a large pool of potential candidates across all the state supreme courts—at least compared with the smaller, more concentrated pools of candidates for judicial promotion in New Zealand and the United Kingdom—a judge’s probability of being promoted to the High Court of Australia is very low.

C. Respect of Colleagues

There is a large law and economics literature that suggests that reputation among one’s peers is a powerful nonpecuniary motivating factor for many people. In this respect, judges are likely to be the same as everyone else, although there is room to debate how wide a judge’s reference group will be. Miceli and Coşgel suggest that “how a judge is viewed by his or her colleagues, by law professors, law students, and the general public” is a motivating factor to perform well. In Australia and New Zealand, where judges are less visible to the public than Justices of the United States Supreme Court, few members of the general public would know of any members of even the highest courts, so public esteem is unlikely to be a motivating factor. For judges of the High Court of Australia and the New Zealand Court of Appeal, whose opinions appear in casebooks and are debated in journal articles, they may well care how their output is viewed by legal academics and law students. In Australia, this might also be true for judges of state supreme courts, who are fairly visible to legal academics, at least those located in the same state.

For most “run of the mill” judges of lower courts, whose opinions are rarely discussed in law lectures or extracted in casebooks, respect of the academic community is unlikely to be a major motivating factor. These judges are likely to have a more narrow reference group, which is restricted to their colleagues in the same court and the lawyers who appear before them. The breadth of the reference group is important when thinking about what motivates judges. For a lower court judge, the main indicator of judicial performance might be how well he or she manages his or her caseload, and esteem among the relevant reference group will be given or withheld on this basis.


40. Miceli & Coşgel, supra note 19, at 32; see also Robert D. Cooter, The Objectives of Private and Public Judges, 41 PUB. CHOICE 107, 129 (1983) (arguing that judges seek prestige among the lawyers and litigants who bring cases before them); Posner, supra note 12, at 13 (arguing that judges care about how they are viewed by lawyers, but not litigants, who appear before them).

art of writing good opinions, which are cited frequently or published in casebooks or legal commentary, is likely to be less important for these judges. Schauer even suggests that if lower court judges spend too much time writing opinions (or worse still journal articles), then this will have a negative effect on how their colleagues perceive them—they will be branded as too ambitious for spending too much time on self-advertising or as free riders because time spent on writing elaborate opinions detracts from time that could be spent on moving the docket along.\footnote{Schauer, supra note 13, at 633.} In contrast, judges of higher (appellate) courts, while also seeking the respect of their colleagues on the bench and practicing profession, have more opportunities to be cited in other opinions and in academic references. These judges will have greater incentive to spend more time writing better opinions.

D. Influence

Posner argues that other factors motivating judges are the power to influence outcomes through voting and the evolution of the law through creating precedent.\footnote{Posner, supra note 12, at 15-23.} Landes and Posner posit that judges derive utility from imposing their policy preferences on the community.\footnote{William M. Landes & Richard A. Posner, Legal Precedent: A Theoretical and Empirical Analysis, 19 J.L. & ECON. 249, 274-75 (1976).} Schauer suggests that “judges could plausibly select outcomes, or select substantive or methodological ‘trademarks,’ for the purpose of maximizing their own influence.”\footnote{Schauer, supra note 13, at 633.} In this respect, the time final appellate court judges spend on writing opinions to increase influence might be a function of the state of the law. It might be argued that, generally, the evolution of the law follows a steady-state growth path. There will be periods, however, when the law is in a state of historical flux. This was the case, for instance, in the High Court of Australia in the 1990s, which is often attributed to the activist legacy of the Mason Court (1987-1995). One commentator on the High Court in the 1990s noted that the law is contestably “more evident and accepted than ever” and that “there are seen to be many more legitimate approaches to argumentation than ever before.”\footnote{Graeme Orr, Verbosity and Richness: Current Trends in the Craft of the High Court, 6 TORTS L.J. 291, 292 (1998).}

There are analogies between the argument being made here and the external environment facing the firm. In the steady-state growth path, the firm will be operating in a technologically stable environment. In most circumstances, decisions will be routine, repetitive, and follow well-defined patterns; opportunities for technological breakthrough and, consequently, for making sizeable economic rents
will be relatively small. In contrast, in periods of flux, there are more opportunities for endogenous innovation—which suggest several possible multiple trajectories where focus is on the creative, rather than allocative, functions of the market—and there is potential for high returns in a high-risk environment.\(^{47}\) In these periods of instability in the evolution of the common law’s long-run growth path, as occurred in Australia in the 1990s, there is greater potential for judges to earn economic rents in the form of influence over the evolution of precedent. In these circumstances, one would expect the judge, as *homo economicus*, to write longer and more complex opinions in an attempt to win the battle of ideas. This is what happened in the High Court in the 1990s; there was a sizeable increase in the number and length of concurring opinions.\(^{48}\) Leaving a legacy in the form of a body of precedent could be an important motivating factor for many final appellate judges to write complex and lengthy opinions. This is more pressing in Australia and New Zealand relative to most jurisdictions in the United States, because in Australia and New Zealand judges face mandatory retirement ages.

The objective of Choi and Gulati’s tournament is to identify highfliers for promotion.\(^{49}\) While the highfliers might have a relatively low marginal time preference for leisure, there is considerable evidence that “run of the mill” judges attempt to minimize workloads. This is most true of lower courts. Cohen shows that the district judges in the United States behave so as to minimize workloads.\(^{50}\) Taha finds that district judges with lighter workloads have more time to publish their decisions.\(^{51}\) However, it is also the case for some judges on appellate courts. Clearly, not all judges are motivated by the prospect of leaving a legacy. Throughout the history of the United States Supreme Court, there have been Justices who appear to have had a preference for minimizing their workload by writing few opinions and free riding on the efforts of others.\(^{52}\) This is also the case for the High Court of Australia, with some Justices well known for hav-


\(^{48}\) For jurimetric evidence on opinion-writing trends over the history of the High Court of Australia, which documents an increase in the length of opinions and number of concurring opinions in the 1990s, see Matthew Groves & Russell Smyth, *A Century of Judicial Style: Changing Patterns in Judgment Writing on the High Court 1903-2001*, 32 Fed. L. Rev. 255 (2004).


\(^{51}\) Taha, * supra* note 36, at 20.

ing a proclivity for joining in joint opinions or writing short concurring opinions.

IV. HOW SHOULD WE MEASURE AND RANK JUDICIAL PERFORMANCE?

A. Opinions Published

In their proposed tournament of judges, Choi and Gulati treat the number of published opinions as an important indicator of judicial productivity. The rationale is that it takes greater effort and skill to write more published opinions. Measuring effort exerted by a judge is an important element of the tournament because comparing the past effort levels of the various judges (a) helps predict future effort levels (we want Justices who will exert high levels of effort) and (b) helps determine who among the lower court judges should be rewarded for their efforts . . . .

On the face of it, this reasoning is sound in the United States context. There is much evidence that writing published opinions involves expending more effort than writing unpublished opinions. In their analysis of published and unpublished opinions by the United States courts of appeals, Reynolds and Richman conclude that “anyone who reads even a small number of unpublished opinions must conclude, given their brevity and informality, that considerable effort has been spared in their preparation.” This reasoning, however, does not extend to Australian and New Zealand courts, which follow the English practice of seriatim opinion writing and use a different reporting system.

In Australia and New Zealand, there is no single signed opinion of the court where the author is readily identifiable. Instead, each judge writes his or her own opinion. Sometimes two or more judges, or more rarely the whole court, give a joint opinion, but who actually wrote the opinion in such situations is not disclosed. Those who endorse the opinion are listed in order of seniority at the beginning. The actual opinion might have been written by one or more of the judges who “sign on.” Attempts to measure judicial output over time are further complicated by changing opinion-writing practices over time. For example, on the High Court of Australia, particularly in its early years, it was common for Justice X to write a short concurring opinion of the form “I agree” with Justice Y, while at other points in time, Justices X and Y would write a joint opinion. In the former

54. Id. at 42 (footnote omitted).
case, the contributions of Justices X and Y are separable products, but in the latter case, it is impossible to know who wrote the opinion.

The system of law-reporting is different in Australia and New Zealand in the sense that in the past, the decision to publish a case was based on its perceived precedent value, rather than the inherent reasoning of the judge(s). Because only a small proportion of decisions were reported in the official and unofficial law reports, published opinions were not a good indicator of effort. In Australia and New Zealand, unpublished opinions have precedent value in the same way as published opinions. Before the widespread use of the Internet, relatively few unpublished opinions were cited because of the difficulty of locating them. However, since the development of the Internet, unpublished decisions are readily available on the Web, free of charge, for a range of courts, and these are widely cited as if they were in the law reports. Indeed, publishing lags, coupled with high purchase costs, are increasingly making law reports redundant. At least at the appellate level, there is no difference in effort needed to produce published opinions that appear in the law reports and unpublished opinions that are available on the Internet.

The practice of seriatim opinion writing, where the author of joint opinions is not disclosed, raises the well-known free-rider problem with team production. The problem with team production, as Alchian and Demsetz note, is that the individual marginal products of team members cannot be verified. The implication is that determining who shirks and by how much is not possible with such a joint product. This comes back to the motivations of judges discussed above. The “run of the mill” judges will have an incentive to shirk by participating in joint opinions to minimize their workload in the absence of effective monitoring. This will be particularly true in the face of increasing workloads. There are instances throughout the history of the High Court of Australia where this is apparent. For example, even casual inspection of the Commonwealth Law Reports suggests that Rich (1913-1950) and McTiernan (1930-1976), the two longest-serving members of the High Court of Australia, frequently free rode on the work ethic and intellect of Dixon (1929-1964), who is Australia’s most celebrated judge, by participating in joint opinions or writ-

---


ing short concurring opinions. Starke (1920-1950), who was a contemporary of all three judges on the Court, caricatured McTiernan as a “parrot” who always copied Dixon. While this characterization was motivated by personal differences, it is supported by studies of voting behavior, which show that McTiernan exhibited a high level of agreement with Dixon. Diaries that Dixon kept provide some insights into the free-riding problem:

Numerous entries show [Dixon] regularly helping Rich with his judgments (even when Dixon had not sat on the case), and occasionally he helped McTiernan with his. Later entries state that he wrote some of Rich’s judgments for him, though it is possible that Dixon meant he was writing sections of Rich’s judgments.

Under these circumstances, it will be difficult to detect the true level of effort judges expend using opinions as a measure of performance. The only real monitoring device will be norms on the court as to whether a judge’s contribution to the workload of the court is acceptable, and this is difficult for those outside the court to effectively observe, let alone quantify in a meaningful manner.

One approach to detecting effort levels, which distinguishes between full written opinions and short concurring opinions, would be to examine the length of the opinion. The argument is that longer opinions address the relevant issues more fully and require more effort. Such a measure attempts to tap into the motives of judges. Judges who want to leave their mark on the law will write longer opinions. Currie uses “pages per year” in constitutional cases to explore the question of who is the most insignificant Justice in the history of the United States Supreme Court. In their study of the determinants of promotion from the English Court of Appeal to the House of Lords, Salzberger and Fenn attempt to confront the free-rider issue by giving short concurring opinions a lower weighting.

There are, however, problems with focusing on the length of opinion. One problem is that it gives a biased indicator of effort expended over time. As discussed earlier, if judges behave as homo economicus, opinions will be longer when there are more opportunities to leave a legacy, which will occur at points in time when the evolution of the

64. Currie, supra note 52, at 469.
65. Salzberger & Fenn, supra note 38, at 837.
law is not on its long-run growth path. Measures such as “pages per year” have to be examined relative to opportunities and in the context of changing norms governing opinion writing over time.\(^6\) Another problem is that, even when examining performance at a given point in time, there is an argument that verbosity should not be rewarded.\(^6\) The increase in the length of opinions on the High Court of Australia in the 1990s, interspersed with increasing numbers of footnotes and citations to academic references, is a trend which has been criticized rather than applauded in some quarters.\(^6\) It has been criticized for increasing the cost to the “consumers” of the opinion, who are defined as those who need to take time to read and digest the implications of the opinion.\(^6\)

Another problem in counting opinions, particularly in the Australian and New Zealand context where there is no effective distinction between published and unpublished opinions, is that there is no product differentiation between publishing outlets—that is, there are no “first-tier,” “second-tier,” and “third-tier” law reports. Studies that attempt to measure the productivity of academic departments or individual academics by counting publications will invariably use some sort of quality-adjusted output measure, so that a publication in a first-tier journal will count for more than a publication in a third-tier journal. An interesting feature of publishing judicial opinions is that at the time of publication, there is either no or little peer assessment of quality. This, of course, is also true of student-edited law reviews in the United States, although there are first-tier, second-tier, and third-tier law reviews; and through the multiple submission process, it is arguable that a manuscript will end up at about its right level.\(^7\) In a Coaseian world, free of transaction costs, one solution might be to have a series of electronic law reports (tier one, tier two, tier three, tier four, and rejected) similar to the Berkeley Electronic Press’s “families” of journals.\(^7\) In practice, though, the transaction costs in-

\(^6\) See Easterbrook, supra note 52, at 491-92 (arguing that “pages per year” has to be examined relative to opportunities and viewed along with other indicators as a measure of judicial performance).

\(^6\) Cf. Brian Leiter, Measuring the Academic Distinction of Law Faculties, 29 J. LEGAL STUD. 451, 467-68 (2000) (arguing that productivity is an important measure for determining the academic distinction of law faculties but that there must be quality checks in place to avoid giving credit for “junk”).


\(^6\) Id. at 64.

\(^7\) Note that this differs from Australia, where student-edited law reviews are peer reviewed.

volved in such a proposal would be high, and it would be complicated by the perceived precedent value of the decision, which may have nothing to do with the effort expended or quality of the writing. There is peer assessment of opinion quality at the time of consumption—if the consumer does not think the publication is worthwhile, it will not be cited. However, this raises the issue of how to determine quality. When citing journals, academics can use the ranking of the journal as a signalling device of quality because highly cited journals tend to be highly read journals. No such signalling device exists for judges. Typically consumers will base their decision on the reputation capital of the individual judge or, perhaps, the circuit.

B. Citations

Choi and Gulati suggest that citations to published opinions are a market test for the quality of opinions.72 Adjusted citation counts to judicial opinions have been widely used in recent years to measure what has variously been described as a judge’s influence, prestige, or reputation.73 These studies generally emphasize that citations are at best a rough proxy for quality, although in the subsequent empirical implementation of their tournament, Choi and Gulati treat citation counts as a measure of the quality of the judicial opinion.74 Citations have also been used in studies of the relationship between aging and judicial productivity, that is, as a measure of the quality of judicial output as a judge gets older.75 Some have criticized the use of citation counts as a measure of quality, focusing on the limitations of their

72. Choi & Gulati, supra note 1, at 305-06.
74. Choi & Gulati, supra note 49, at 48-61.
use. One such limitation is evidence of like-minded citations where "conservative" or "liberal" judges tend to cite members of their respective group to the exclusion of the other camp. This is problematic where the court is dominated by appointments on one side of the political fence. This has traditionally been the case in Australia and New Zealand, where most appointments to the High Court of Australia and New Zealand Court of Appeal have been made by the "conservative" side of politics. The problem in Australia, at least, has been exacerbated by the unorthodox approach to opinion writing by some Labor appointments, which have been shunned by their more "conservative" colleagues. While this issue and problems with citation counts are well known, it is generally recognized in a wide range of disciplines that in spite of these limitations, there is a correlation between citations received and the quality of output.

One problem with using citation counts to judicial opinions in the United States is the problem of ghostwriting, where it is the law clerk, rather than the judge, who writes the opinion. As Landes, Lessig, and Solimine note, "[c]itations may reflect the influence or quality of law clerks as 'ghost writers' rather than that of judges . . . ." However, as Choi and Gulati argue, the problem is tempered because judges who rely heavily on clerks will produce opinions of more varying quality, because it will be difficult even for a high-quality judge to consistently retain high-quality clerks. Moreover, even with judges who rely heavily on law clerks, citations might still reflect performance as a manager. The reasoning is that there will be a signalling device where the best clerks (who write the best opinions and get cited the most) will be attracted to the best judges, although the market for clerks is imperfect. The issue of ghostwriting does not impede using citation counts in Australia. The limited anecdotal evidence that exists from both judges and former clerks suggests

77. Austin, supra note 76, at 833-34.
78. See Choi & Gulati, supra note 49, at 48 n.38 (citing a multitude of studies "that either suggest or assume a relationship between citation counts and quality").
80. Landes et al., supra note 73, at 274.
81. Choi & Gulati, supra note 49, at 52.
82. Cass, supra note 13, at 989-91.
83. Epstein, supra note 11, at 841-44.
that judges write their own opinions in Australia. There is, however, another practical problem with using citations to measure the quality of judicial opinions in Australia and New Zealand. This stems from the earlier observation that there is no signed opinion of the court. Thus, when a case is cited, it is impossible in most instances to attribute authorship. Therefore, existing studies of judicial influence and productivity for Australian courts have relied on invocation rates. 86 Choi and Gulati suggest that invocation rates are more an indication of reputation, while citation counts are a measure of quality. 87 It can be argued that invocation rates are preferable to citation counts as a measure of the influence of the opinion on the citing judge’s thinking, even in the United States, because there is no requirement that the citing judge go the extra step and refer to the author of the cited opinion by name. 88

C. Reversals

Posner suggests that while judges do not like to be reversed, aversion to reversal does not figure prominently in the judicial utility function. He reasons that most reversals reflect differences in judicial philosophy or legal policy, rather than error by the judge below. 89 However, while some reversals will undoubtedly reflect policy issues, it is reasonable to assume that a judge gets disutility when his or her decision is reversed by judges who are in higher courts or by future judges. 90 Elsewhere Posner accepts that since judges might seek to be promoted, they may display extreme “sensitivity to being reversed.” 91 Cass agrees: “Judicial reversal reflects professional criticism by other professionals. It will likely be a negative for almost all judges in almost all circumstances.” 92 For this reason, judges will expend effort on ensuring that their opinions are well-reasoned to reduce the occurrence of reversal. The proportion of cases in which a judge has been reversed is something that is readily quantifiable as one component of a tournament of judges.

86. Bhattacharya & Smyth, Aging and Productivity and the High Court, supra note 75, at 203; Bhattacharya & Smyth, Judicial Prestige and the High Court, supra note 73, at 233-34; Smyth & Bhattacharya, Aging and Productivity and the Federal Court, supra note 75, at 149-50; Smyth & Bhattacharya, Judicial Prestige and the Federal Court, supra note 73, at 238-39.
88. Klein & Morrisroe, supra note 73, at 375-76.
92. Cass, supra note 13, at 984.
Previous studies for a range of courts both inside and outside the United States have used reversals as a measure of judicial performance. Posner uses reversals by the Supreme Court to assess the quality of the United States Court of Appeals for the Ninth Circuit relative to other circuits. Reversals have also been employed in studies of the English Court of Appeal and the New Zealand High Court as an indicator of the quality of decisionmaking. While Maitra and Smyth found that reversals had a statistically insignificant effect on promotion prospects of judges on the New Zealand High Court, Salzberger and Fenn found that those judges on the English Court of Appeal who had a higher proportion of their decisions reversed by the House of Lords were less likely to be promoted.

D. Judicial Gowns or Academic Robes?

Judges’ contributions to academic scholarship in the form of law review articles has been described as a form of judicial leadership. If so, it is arguable that extrajudicial contributions such as academic writing should be a part of the tournament. Currently, if anything, judges in the United States vying for appointment to the Supreme Court have a disincentive to publish in law reviews, because it contributes to a Bork-like paper trail of their views. There is empirical support for this view. Gaille examined the academic publication patterns of judges on the United States courts of appeals prior and subsequent to the Bork confirmation hearings and found that judges published less in academic journals after the hearings.

The main argument supporting the view that academic publications should be viewed favorably is that there are complementarities between good judging and good writing abilities reflected in law review articles. Thus judges who are novel thinkers are more likely to publish in law reviews, and this proclivity is correlated with citations received. On the United States courts of appeals, academics turned judges, such as Easterbrook and Posner, are cases in point.

94. Id.
95. E.g., Maitra & Smyth, supra note 38, at 221; Salzberger & Fenn, supra note 38, at 837.
97. Salzberger & Fenn, supra note 38, at 842.
100. Klein & Morrisroe, supra note 73, at 383.
101. See Choi & Gulati, supra note 49, at 74 (finding that Easterbrook and Posner come out at the top in the tournament of judges).
ever, in these cases, it is their citation counts, rather than scholarly publication records, that make them good judicial performers. There is an alternative argument that outsiders on the court are more likely to publish in law reviews because their views are not gaining acceptance on the court. There is casual evidence to support this on the current High Court of Australia: Justice Kirby has the highest dissent rate in the history of the Court and is also the most prolific publisher in law reviews, publishing three times more than any other Justice. If judges behave as *homo economicus*, then including outside activities, such as writing journal articles, in the tournament of judges also potentially creates perverse incentives. As Choi and Gulati note, “Giving a judge credit for doing other things will diminish the incentive to spend time on opinion writing.”

V. IS IT APPROPRIATE TO MEASURE AND RANK JUDICIAL PERFORMANCE?

Choi and Gulati’s whole rationale for a tournament of judges premised on objective quantitative indicators is that it would make the judicial appointment process more transparent and less politicized. Some, however, have expressed reservations about the extent to which judicial performance can be quantified. The outcome of a tournament of judges has to be viewed subject to these reservations, which fall into two main categories. The first concerns the merits of applying empirical techniques to the study of judicial behavior because of inherent data problems. The second is that quantitative studies cannot capture important qualitative dimensions associated with good judging, and in particular, quantitative studies are unable to measure the positive effects of collegiality. In this Part I, provide a brief review of these concerns.

The law and economics movement, together with allied researchers in political science, has been at the forefront of explaining judicial behavior using quantitative tools. One argument against the expansion of economics beyond its traditional domain of markets is that the data is too thin to reliably test hypotheses formulated in the social sciences. While markets generate a range of data on variables

102. *Id.* at 68 tbl.11.
106. *Id.* at 29-30.
of interest—such as output, prices, and employment—many studies of courts and judges are restricted to a universe of published, written decisions. The focus on such decisions, along with an emphasis on appellate courts, reflects greater data availability, reduces the extent to which findings can be generalized, and raises the prospect of selection bias. In some studies of judicial behavior, sample size has been problematic, with far-reaching conclusions drawn on the basis of a small number of observations.

In studies of judicial decisionmaking, there are also problems with comparability and experimental design. Prior to the 1980s, most empirical work on judicial decisionmaking used discrete datasets bearing on a limited set of research issues, which made it difficult to compare and replicate results because different studies often made different assumptions or collected data on variables in different ways, or both. Longitudinal studies, employing time series data over time, have often had to make the heroic assumption that cases from different eras are sufficiently similar that the results are comparable and generalizable. One approach to addressing the limitations on the assumption of comparability has been to explore alternative research designs such as experimental methods. The problem, though, with experimental studies that employ simulated cases to gauge judicial performance is the associated loss of authenticity.

Having said this, empirical researchers aware of the limitations have recently made considerable progress dealing with many of these data issues. One development has been a trend in the United States toward the use of datasets containing both published and unpublished decisions in recognition of the biases introduced with only using published decisions. A second occurrence in the United States has been the emergence of multiuser datasets designed for use in a range of research problems. These datasets are not restricted to the Supreme Court but have also been developed for the courts of appeals and state supreme courts. A third development, which is particularly apt in the context of this Essay’s Antipodean focus, is that studies looking at different aspects of judicial performance are starting to emerge for courts outside the United States. It has to be said,

109. The rest of this paragraph draws on Heise, supra note 13, at 845-48.
111. Heise, supra note 13, at 846.
112. Id. at 830.
113. See id. at 848-49.
though, that progress on this front has been patchy. Atkins’ observations, made at the beginning of the 1990s, are still true now: “Despite much progress during the past few decades in the study of courts and judicial behavior, most theory and data developed by social scientists for understanding legal systems still remain very much the product of, and thus bound to, the inevitable peculiarities of the U.S. context.”

Atkins goes on to suggest, “This must represent a curious inconsistency in the logic of our research agenda, since the scientific method to which we adhere is supposed to encourage theory building and description at the broadest level possible.” These comments underpin the need, emphasized in the previous Part, to be flexible in the design of a tournament of judges intended to measure judicial performance and to cater to different institutional contexts.

Another objection to attempting to quantify judicial performance is the argument that quantitative tools are not apposite to capture the multifaceted aspects of being a good judge. Hensler, for instance, argues, “Researchers simply do not have available very good quantitative approaches to studying large social organizations [such as courts] or interaction processes [within courts].” The argument is that there are important qualitative aspects to being a good judge which are difficult to quantify in a tournament. In addition to diligence and ability, which can be measured using opinion and citation counts, other generally agreed useful qualities include integrity, fairness, temperament, and collegiality. Judge Edwards has been one of the strongest critics of attempts to quantify judicial performance on the basis that collegiality cannot be measured:

Collegiality is a qualitative variable in appellate decision making, because it involves mostly private personal interactions that are not readily susceptible to empirical study. Regression analysis does not do well in capturing the nuances of human personalities and relationships, so empirical studies on judicial decision making that rely solely on this tool are inherently flawed.

Some of these desirable qualities, though, have been captured in surveys such as that of the American Bar Association. Moreover, contrary to what the quote from Judge Edwards implies, there is much

---

115. Id.
evidence of the success of regression analysis in capturing these nuances of human behavior. For instance, there have been hundreds of published articles in which regression has been used to analyze scores from personality inventories, such as the Minnesota Multiphasic Personality Inventory, the 16PF, and the Myers-Briggs Type Indicator, where regression analysis has been used precisely for this purpose.120

Having said this, Edwards’ concerns have important implications for how the efficiency aspects of measuring judicial performance are interpreted. The Alchian and Demsetz team production model, with its emphasis on monitoring the individual, is concerned with allocative efficiency.121 In the Alchian and Demsetz model, the firm and the market are opposite sides of the same coin.122 In the context of this discussion, “the firm” can be interpreted as the court. In the Alchian and Demsetz conception, there is no role for the court as a social organization. The efficiency gains stem from specialization and the detailed division of labor. But the emphasis Judge Edwards and others place on the importance of collegiality in the court suggests that in the court’s team production, there are potential benefits from organizational learning. These stem from continual interaction between members of the court that cannot be captured with a more narrow focus on allocative efficiency. There are positive externalities in team production from collegiality that are manifest in the form of collective learning on the court. The relevant efficiency criteria is productive efficiency rather than allocative efficiency, and the efficiency gains accrue from economies of scope rather than economies of scale.123 Edwards’ concerns suggest that attempts to measure judicial performance using similar tournament designs to Choi and Gulati’s focus heavily on the allocative-efficiency dimension. The productive-efficiency dimension tends to be downplayed. Thus, it is important to recognize that the hardheaded figures which come out of a tournament based on citation counts, while insightful, tell only part of the story.

A different issue is whether the acceptance of broader aspects of human behavior—such as collegiality, integrity, and fairness as desirable attributes in a judge—is commensurate with a conceptualization of the one-dimensional homo economicus. One might argue that

120. For a meta-analysis, see James M. Schuerger et al., Factors That Influence the Temporal Stability of Personality by Questionnaire, 56 J. PERSONALITY & SOC. PSYCHOL. 777 (1989).
121. See Alchian & Demsetz, supra note 58, at 779-81.
122. Id. at 789-94.
123. For an extended theoretical discussion of these ideas in a different context, see Russell Smyth & Dic Lo, Theories of the Firm and the Relationship Between Different Perspectives on the Division of Labour, 12 REV. POL. ECON. 333 (2000).
the model on which he is conceived is at best myopic and at worst incommensurable with regard to the component of perceived moral obligation that arguably influences much of human social behavior, including behavioral intentions and, importantly, attitudes. If this is the case, “Justice Homo Economicus” might be a distorted lens through which to view judicial behavior. Instead, perhaps we should be asking what motivates “Justice Homo Sociologicus,” the (irrational?) mirror image of economic man. The teleological underpinnings of the two judges differ in that homo economicus has essentially an ego-goal orientation, whereas homo sociologicus is a role player acting under the instructions of his environment. While homo economicus is assumed to be motivated by self-interest, homo sociologicus is conceived in the constructivist tradition, which acknowledges that all actors are resident within some social context. The social context, or more precisely the social norms that emerge from within the social context (including norms governing judges’ interaction on the court), influence behavior and, important in the context of measuring judicial performance, the pursuit of goals. This is because they impose upon judges social roles, inherent in which are prescriptions about behavior, group interaction, and moral responsibilities.

Another challenging aspect of measuring and ranking judicial performance centers on how to choose the most appropriate performance standard referents. Perhaps due to the fact that we have well-developed quantitative methods to assess individual performance relative to group performance, performance measurement has traditionally (and it seems stubbornly) relied upon extrinsic normative referents. Reference to extrinsic norms lends itself nicely to performance ranking. This is useful, because a ranking is a clear and concise manner in which to understand performance. Ranks are conceptually simple, and they are not readily misinterpretable; so in these terms, adopting a ranking approach maximizes the golden rules of measurement: validity and reliability. Having said that, it seems strange to suggest that these normatively referenced performance measurement “habits” that we have gotten into might indeed not be ideal when it comes to a tournament of judges. But consider this: While one’s performance ranking within a group is undoubtedly informative, it is limited, because assessing performance relative to a group norm and to observed group performance bounds assumes a commonality between the observed upper performance bound of the group and the potential upper performance bound of the highest flier. But what if the highest flier is in fact underperforming in terms of

his or her own potential performance bound? What if he or she could do better, but simply does not need to in order to maintain a number one ranking?

This brings us fundamentally to the raison d’être of a tournament of judges. One fundamental purpose of a tournament of judges indeed should be to identify those ripe for promotion—and these are likely the normatively referenced highfliers. Another purpose, of course, should be to identify those who are dragging their heels. But the assumption that the highfliers never drag their heels is a dangerous one. Certainly in competitive sports, it is entirely conceivable that some elite athletes could perform well below their personal best yet claim a gold medal. It should be no different within other domains. Hence, in terms of the performance referent “sensibility,” it seems imperative to look beyond extrinsic referents and to also consider the assessment of performance in terms of self-referenced potential performance parameters.

While such traditional nomothetic methods as the use of Z-score-type analysis tell part of the performance story—and probably tell it well for “run of the mill” judges—a tournament of judges, in which such methods are to be exclusively adopted as the tool to identify the shining and not-so-shining stars, seems to require a set of tools not to be found in the typical bag of tricks of the nomothetic research analyst. One way of measuring performance within a tournament of judges that attempts to tackle these issues may be to adopt the “idiographic” approach developed by Lamiell.125 This approach, developed and utilized for the most part in the psychology literature pertaining to personality, represents a blending of the idiographic and nomothetic research paradigms. Unlike either pure idiography, with its case study methods and focus on the qualitative understanding of development and performance, or pure nomothesis, with its aggregate statistical methods and focus on the group, Lamiell’s idiothesis provides a vehicle for the marriage of these seemingly incommensurable research positions.126 Lamiell has presented a statistical model for the study of individual differences in performance that shifts the study referent from some measure of central tendency to the self.127 In simple terms, the model assesses performance not in terms of what others do, but in terms of what a person tends not to do but


126. See James W. Grice, Bridging the Idiographic-Nomothetic Divide in Ratings of Self and Others on the Big Five, 72 J. PERSONALITY 203, 204-05 (2004); Albert Silverstein, An Aristotelian Resolution of the Idiographic Versus Nomothetic Tension, 43 AM. PSYCHOLOGIST 425 (1988).

127. See Lamiell, Toward an Idiothetic Psychology, supra note 125, at 281-86.
could do. In presenting this model, two main advances are made: first, the individual differences “error” variation that is suppressed in traditional inferential statistical models becomes friend, and not foe; second, the positivist scientific rigor that is “compromised” using idiography is retained. In terms of a tool for assessing performance within a tournament of judges, this approach has several advantages. First, it can be applied to populations as small as \( N = 1 \), and hence the naturally occurring small population of judges is no longer a methodological stumbling block. Second, it does not place unrealistic performance expectations or impose unfair value judgments upon those judges who will never be highfliers. Third, it does not allow the highfliers to stop too often to smell the roses, live off their reputation, or do any of the things that past successes might tempt us all to do every once in a while.

The challenge, of course, is how we determine what one could do; or how we construct an index to estimate the bounds of individual performance expectations. This is an area that has received scant scholarly attention and, to my knowledge, none within the area of judicial performance. Certainly the magic answer is not to be found in this Essay; but my purpose will be served if it stirs the reader in any way to consider the possibilities that such a model may present for better measurement of judicial performance. Any index of what one could do needs necessarily to account for both individual differences and accepted norms of what the peer group can do and has done. Yes, there is certainly a place for extrinsically referenced norms within idiothesis. After all, one only gets to be a judge because many years before one scored high enough relative to one’s peers to attend law school, and so on. Thus current measures of judicial performance discussed earlier in this Essay probably have a place in the model, as does past performance, since generally there is a high positive correlation between one’s past and future performance. But, these should be assessed relative to individual differences in the cognitive, motivational, and social peculiarities of each judge that affect day-to-day functioning, as well as relative to peer norms. Again, while research is scant in terms of applications to judges, there is a

\[128. \text{ Id. at 281.} \]
\[129. \text{ See id. at 281-86.} \]
\[130. \text{ For an application of the idiothetic approach to vocational choices, see Terence J.G. Tracey & Maria Darcy, An Idiothetic Examination of Vocational Interests and Their Relation to Career Decidedness, 49 J. COUNSELING PSYCHOL. 420 (2002). For an application of the idiothetic approach to exploring the competence of counselors in multicultural settings, see Maria Darcy et al., Complementary Approaches to Individual Differences Using Paired Comparisons and Multidimensional Scaling: Applications to Multicultural Counseling Competence, 51 J. COUNSELING PSYCHOL. 139 (2004).} \]
\[131. \text{ See Icek Ajzen, Residual Effects of Past on Later Behavior: Habituation and Reasoned Action Perspectives, 6 PERSONALITY & SOC. PSYCHOL. REV. 107, 107-08 (2002) (stating that past behavior is a good predictor of future behavior).} \]
large body of literature that demonstrates that occupational performance is affected by such things as individual differences in motivation and personality, as well as external life events. Each of these contextual factors will bear upon what a judge could do. And each of them will be different for different judges.

Finally, if one were able to construct such an index, how would one use it as a decisionmaking tool within a tournament of judges? In terms of identifying those who appear to be underperforming, the index would be useful in determining whether intervention would be likely to address that underperformance. If underperformance is identified relative to a self-referenced performance bound, then it may be that manipulation of a contextual factor—say, motivation—would improve performance. Once improved, this judge may be an ideal candidate for promotion. But if underperformance is identified relative to a group-referenced performance bound, and if contextual factors cannot explain the underperformance, then it simply may be that current performance levels are being maximized. This person is clearly not one bound for promotion. The high flier who excels relative to both the group and self-referents may not be the ideal candidate for promotion in the long term. While there will always be those who will excel normatively without maximum effort, they are few and far between. It may well be that the judge who performs consistently at a self-referenced best will be the one whose career is cut short by occupational burnout or one of the many stress-related diseases well demonstrated to be associated with the Type A “workaholic” personality. Maybe the judge most deserved of promotional attention is the one who is a good performer relative to the group, but slightly less so in relation to the self. Clearly this judge has the intellectual wherewithal necessary for promotion. The key to understanding this person is to understand what is holding him or her


134. See Bruce D. Kirkcaldy et al., The Influence of Type A Behaviour and Locus of Control upon Job Satisfaction and Occupational Health, 33 PERSONALITY & INDIVIDUAL DIFFERENCES 1361, 1367-70 (2002). For a Singapore study of the causes of work stress among lawyers and five other professional groups, with a particular focus on the effects of personality on work stress, see Kwok Bun Chan et al., Work Stress Among Six Professional Groups: The Singapore Experience, 50 SOC. SCI & MED. 1415 (2000).
back. Perhaps he or she is bored and a new and more fulfilling role will spark his or her occupational enthusiasm. Perhaps he or she is experiencing non-work-related problems—the impact of which may well dissipate with intervention. There are many “perhapses,” and the discussion of the many possibilities is beyond the scope of this Essay. The challenge is to concentrate research efforts on identifying the most salient “perhapses” for judges. Once we have a better understanding of these factors, our task in constructing the most parsimonious idiothetic performance model will be all the easier.

VI. INCENTIVE EFFECTS OF MEASURING AND RANKING JUDICIAL PERFORMANCE

Choi and Gulati emphasise the positive incentive effects of their proposed tournament:

We . . . believe that a judicial tournament would provide appellate judges with the otherwise absent external incentive to exert greater effort than they currently do. If high effort as a circuit judge (for example, publishing more opinions or hearing oral argument in more cases) is a criterion for promotion, rewarding effort with a higher ranking would induce circuit judges to work harder at their jobs.135

If, in fact, judges do behave as *homo economicus*, it is reasonable to expect that a tournament with the elements proposed by Choi and Gulati should generate positive incentive effects, at least for appellate judges whose opinions are read and cited. At the very least, a high objective ranking in such an exercise brings professional respect and increased visibility, which in turn contribute to the prospects of promotion. Choi and Gulati make the good point that while judges sitting on the same court probably already have a good sense of each other’s abilities, objective rankings can have positive incentive effects even on mid-level judges with little chance of promotion, because they can use the rankings to market themselves to those outside the court, such as potential law clerks.136

There is, however, a problem. Choi and Gulati’s application of their tournament137 produces such interesting results, which are useful in providing objective criteria in relation to who would make a good Supreme Court Justice, precisely because the rules of the tournament were not preannounced. If the President were to announce that from now on Supreme Court Justices would be appointed based

136. *Id.* at 314; see also Patricia M. Wald, *The Rhetoric of Results and the Results of Rhetoric: Judicial Writings*, 62 U. Chi. L. Rev. 1371, 1372 (1995) (arguing that one reason judges publish their opinions is to attract the best law clerks).
on the outcome of a Choi and Gulati-type tournament of judges, because of the very fact that judges do behave as *homo economicus*, it would have both positive and negative effects on judicial behavior. Williamson describes the new *homo economicus* as someone who engages in “self-interest seeking with guile.”\(^{138}\) This brings us back to the point made in Part II that *homo economicus* will behave in an opportunistic manner to achieve his self-interested objectives.

One potential adverse implication concerns the negative effect of increased competition. Choi and Gulati suggest that one of the positive spin-offs of their tournament is that it will provide incentives to compete and that this will result in higher levels of effort.\(^{139}\) In the spirit of the Alchian and Demsetz model of team production,\(^{140}\) the focus here is clearly on reducing shirking and improving allocative efficiency on the court. This, however, might come at the cost of reduced productive efficiency if a more competitive environment has a negative effect on the level of collaboration and potential for collective learning on the court. Chen shows that participants in tournaments have incentives to sabotage others in a bid to improve their own relative position.\(^{141}\) Choi and Gulati seek to downplay this possibility, suggesting that judges are unlikely to undermine their colleagues.\(^{142}\)

But, if one accepts that judges behave as *homo economicus*, there is no reason to think that judges would behave differently from others interacting in small social groups. There is plenty of anecdotal evidence to suggest that courts, like other social organizations, have been affected by personality differences, which a tournament that promoted competition may exacerbate. The High Court of Australia for much of its history has been impaired by personal tensions and petty jealousies between the judges that have adversely affected its work.\(^{143}\) One (albeit extreme) example was Isaacs, who was a Justice of the Court from 1906-1931 and Chief Justice from 1930-1931.\(^{144}\) His biographer notes that he would hide cases and play down the significance of issues in argument so as to give himself a head start relative to the other judges when it came to writing opinions.\(^{145}\)

Making the number of published opinions the criteria for measuring good performance can also have negative side effects. If the focus


\(^{139}\) Choi & Gulati, *supra* note 1, at 313-15.

\(^{140}\) Alchian & Demsetz, *supra* note 58, at 779-81.


\(^{142}\) Choi & Gulati, *supra* note 1, at 309.


\(^{144}\) *ZELMAN COWEN, ISAAC ISAACS* 113 (1967)

\(^{145}\) *Id.* at 124-25.
is on counting opinion numbers, there is always the potential, in the United States setting, that it will promote a publish-or-perish mentality, where the number of published opinions increases but the overall quality of those opinions decreases. One might argue that a judge will not submit “substandard” opinions for publication, because publishing poor-quality opinions would hurt his or her reputation. But this assumes a certain degree of farsightedness, because the reputation effect is likely to occur only with a lag, while in practice judges confronted with the prospect of improving their finishing position in the next tournament may well be myopic. Even if the suggestion made earlier of having multitiered law reports containing opinions of different quality was considered fanciful, a more realistic solution to this problem would be to set the initial hurdle for publication higher.

A related problem with focusing on opinion counts is the potential adverse incentives this has for managing caseloads. Even in higher courts, where the opinions of judges are read, opinions published might not be the best measure of productivity if time spent on crafting elaborate opinions is at the expense of case management. One of the main reasons for criticism of longer opinions in the High Court of Australia since the 1990s has been the view that the time taken to write longer opinions affects the capacity of judges to deal with their overall caseloads within a reasonable period of time. The argument is that longer, multiple opinions which contain excessive citation are responsible for an increase in the average time elapsed “between the conclusion of [the] hearing and pronouncement of a judgment which has been reserved.” This is a sensitive issue in Australia, as in the United States, where increasing backlogs have intensified community pressure for speedier justice. In 1998, court delays came under increased media scrutiny in Australia, when New South Wales Supreme Court Justice Vince Bruce faced a State Parliament vote to remove him from office for taking up to three years to deliver a judgment. This has forced courts to monitor delays. In an attempt to address this issue, judges of the New South Wales Supreme Court are now required to make monthly progress reports on delivery of their judgments and time limits have been placed on the time period

---

146. See supra notes 70-71 and accompanying text.
148. Campbell, supra note 68, at 63; see also Doyle, supra note 147, at 738.
between committal hearings and the trial in criminal cases. Similar procedures operate in the United States. For example, on the United States Court of Appeals for the District of Columbia Circuit, a judge who has three or more assigned opinions pending that are not in circulation is not allowed to sit on any new cases until the backlog is cleared.

One might argue that the approach to avoid the problem of judges publishing large numbers of poor-quality decisions is to put more emphasis on citation counts. The reasoning is that poor-quality decisions will not get cited, so the marginal return from expending effort on publishing poor-quality decisions will be low. But if the focus is solely on citation counts, this also presents problems. Kobayashi and Lott suggest that judges who want to maximize citation counts to their own decisions will have an incentive to replace existing precedents with their own inefficient decisions, which will generate more litigation and hence more opportunities to be cited. While Posner argues that most judges would not behave in this fashion, it is reasonable to expect that judges confronted with a tournament that explicitly rewards citation counts and “self-interest seeking with guile” would implement strategies to maximize their citation count. One possible strategy is the formation of citation clubs, where members cite each other’s opinions to the exclusion of others on the court. To some extent this exists already. Studies for the United States courts of appeals and Federal Court of Australia have found that judges who attended elite law schools receive more citations. This might be because the judges from elite schools are more able, or it might be because like-minded judges cite each other. Landes, Lessig, and Solimine prefer the latter explanation, suggesting that on the United States courts of appeals, graduates from Harvard and Yale form “part of [a] large network of ‘like-thinking’ judges who will tend to cite each other more often than they cite other judges.” Choi and Gulati, while acknowledging that judges might cite similar-minded colleagues who agree to reciprocate, argue that this might in fact hurt a judge’s prospect of winning the tournament, because the judge will not be citing the best opinions available; therefore, the quality of the output will fall. But, this is only true if the probability of being

151. Id.
152. Edwards, supra note 14, at 1665.
154. See id.
155. Landes et al., supra note 73, at 324.
157. Landes et al., supra note 73, at 324.
158. Choi & Gulati, supra note 1, at 311 & n.31.
cited is purely a function of the quality of the opinion, which need not be the case in the presence of citation clubs.

There are parallels here with the “insider”-“outsider” distinction in legal scholarship; that is, the argument that while “outsiders”—critical legal theorists, feminists, and critical race theorists—are producing good scholarship, it is not being cited by “insiders.” Levit suggests that, as a result, “[r]eliance on quantitative assessments of legal scholarship may tend to subtly perpetuate existing hierarchies of race, gender, and theory prominence, while telling little about the substantive or foundational qualities of a theory.” There is no evidence of gender or racial discrimination in citation counts on the courts, but the point about theory building is arguably true for citations to judicial opinions, with “outsiders” holding nonmainstream views or employing unorthodox opinion-writing techniques being overlooked. An illustration of an “outsider” on the High Court of Australia was Lionel Murphy, who was a Justice from 1975-1986. Murphy largely ignored the doctrine of precedent, arguing from first principles. At the time he was often in dissent, but many of his ideas have since gained acceptance in the Court’s jurisprudence, albeit in more complex forms. Murphy, however, is rarely cited as being the originator of those ideas.

VII. CONCLUSION

The concept of a tournament of judges is a useful one. In the United States, where the judicial appointment process has become so politicized, and even in the Antipodes, where judicial appointments are not as politicized but still nontransparent, the use of objective indicators of performance would make a positive contribution to the selection process. However, as Choi and Gulati freely admit, a tournament of judges is not the be-all and end-all when it comes to deciding who are the best performers. A potential problem with explicit reliance on a tournament, which this Essay has emphasized, is that it

161. See, e.g., Landes et al., supra note 73, at 324 (finding no evidence of racial or sexual discrimination in citation practices on the United States courts of appeals).
163. Id. at 254-58.
164. See JUSTICE LIONEL MURPHY: INFLUENTIAL OR MERELY PRESCIENT? passim (Michael Coper & George Williams eds., 1997); Kirby, supra note 162, at 254-55.
165. Choi & Gulati, supra note 1, at 322.
can result in game playing.\textsuperscript{166} But, as Choi and Gulati state, “Our primary goal is not to produce winners, but to enable transparency in the nomination process.”\textsuperscript{167} Given that each of the quantitative indicators have positive and negative points, it is important to use a range of quantitative indicators, as Choi and Gulati do in their “composite measures,”\textsuperscript{168} in order to obviate the limitations of any one measure.

While the tools exist to evaluate qualitative dimensions of judicial performance, research in this area is less advanced due, at least in part, to more restricted data availability relative to counting citations and opinions. One approach discussed in this Essay to tackle the qualitative dimension of judging, which has not been utilized to this point, is to borrow idiothetic methods from psychology. It is an important step forward, both conceptually and methodologically, to further refine idiothetic methods that may enable us to move beyond the typically accepted tradition of “extrinsic normative comparison” that persists in performance evaluation. Such comparison, which lends itself to conclusions based upon ranking procedures, seems myopic with respect to identification of negative change among high-level performers and positive change among low-level performers as well as over acceptance of the status quo. In the end, subject to these reservations, the most appropriate role for the tournament proposed by Choi and Gulati might be to help generate a pool of qualified candidates, from which candidates for promotion can be selected. Finally, this Essay has discussed the tournament in the specific context of the Australian and New Zealand legal setting, where several institutional arrangements differ from the United States. These different institutional arrangements do not mean that a tournament of judges has no relevance to the Antipodes. On the contrary, I have argued quite the opposite, but what they do highlight is that the design of a tournament, in whichever court it is played, needs to be tailored to meet the specific institutional context in which the court operates.

\textsuperscript{166} For further discussion of the game-playing issue in the context of designing optimal incentives, see George Baker et al., \textit{Subjective Performance Measures in Optimal Incentive Contracts}, 109 Q.J. Econ. 1125 (1994).

\textsuperscript{167} Choi & Gulati, supra note 49, at 81.

\textsuperscript{168} Id. at 70-75.
I. INTRODUCTION

Many people, it seems, are concerned throughout their lives in varying ways with how others think about or are affected by them—that is, their status, prestige, influence, or reputation. Similar judgments are ubiquitous in our legal culture. They often guide a student’s choice of law school, a lawyer’s choice of firm, area of practice, or which judge to clerk for, or a client’s choice of attorney, to name just a few. They also often guide our views about individual judges, whether on the U.S. Supreme Court or lower courts, both within a particular period and across time.

Status and reputation also guide our view of multimember courts. For example, we often speak, if only indirectly, of differences in reputation of the U.S. courts of appeals\(^5\) or of the U.S. Supreme Court at different periods in its history.\(^6\) We have a sense of some courts dominating others, either in quality, importance, or by other comparative measures. Whether speaking of individual judges or the courts on which they sit, studies of quality immediately run into measurement problems. Status, reputation, and similar concepts seem largely subjective. One way to deal with the problem is to employ various types of citation analysis. Determining how often a particular judge or court has been cited by others can potentially draw more objective differences between judges and courts.

In the past decade there has been a boomlet of studies using citation analysis to gauge the reputation or influence of particular judges.\(^7\) Until recently, however, there has been little systematic study of the reputation or influence of multimember courts as such. My goal in this Essay is to fill some of that gap by exploring the reputations—both historically and at the present time—of the individual U.S. courts of appeals. That is, I will compare and contrast the reputations of the thirteen (as the number stands now) courts of appeals: the First through Eleventh Circuits, the Court of Appeals for the District of Columbia Circuit, and the Federal Circuit.\(^8\)

This Essay proceeds as follows. Part II addresses and disentangles the concepts of reputation, prestige, and influence, particularly as they are used in the legal community. It considers various measures of those concepts, principally, though not only, through citation analysis. I address the pros and cons of that, and other measures, and I explore problems associated with attributing reputation to collective entities, like multijudge courts. In Part III, I address efforts,

---

5. President Clinton, for example, when announcing the Supreme Court nomination of Judge Ruth Bader Ginsburg, of the U.S. Court of Appeals for the District of Columbia Circuit, referred to her service on “the second-highest court in our country.” Transcript of President’s Announcement and Judge Ginsburg’s Remarks, N.Y. TIMES, June 15, 1993, at A24. For a more recent example, see Chris Mooney, Circuit Breaker, AM. PROSPECT, Spring 2003, at A14 (referring to the D.C. Circuit as “the second most powerful court in the United States”).


7. Various types of citation analysis have been used for decades in the legal community to gauge the impact of books, law review articles, court decisions, or judges, among other things. See generally RICHARD A. POSNER, FRONTIERS OF LEGAL THEORY 411-40 (2001); Symposium, Interpreting Legal Citations, 29 J. LEGAL STUD. 317 (2000).

8. Officially, each of the courts I am addressing is “known as the United States Court of Appeals for the [particular] circuit.” 28 U.S.C. § 43(a) (2000). Nonetheless, in this Essay I will, for convenience, use the terms “courts of appeals” and “circuits” or “circuit courts” interchangeably.
both historical and contemporary, to measure the reputations of the courts of appeals. Initially, I consider various accounts from the popular press and other nonscholarly sources that attempt to rank the circuits. Then I turn to somewhat more objective measures, such as surveys of attorneys and federal judges, of the reputations of circuits. Finally, I consider how various studies of citation analysis of the influence of particular federal appellate judges can be brought to bear on the reputation of the circuits on which they sit.

As Part III outlines, historically and to some extent to the present day, the Second and D.C. Circuits have been regarded as enjoying the best reputation and most influence. Yet some, though not all, of the putatively objective measures of influence place the Seventh Circuit far above those two. In Part IV, I explore this apparent disconnect and conclude that a variety of factors have led, in general, to the overall homogenization of reputation among the circuits, with the notable apparent exceptions of the D.C., Second, and especially the Seventh Circuits. In the conclusion to the Essay, I briefly address whether reputation in this context remains a meaningful concept, offer suggestions for future research, and link measures of circuit reputation to the measurement of judicial performance in general.

II. MEASURING JUDICIAL REPUTATION, PRESTIGE, AND INFLUENCE: INDIVIDUAL JUDGES AND MULTIMEMBER COURTS

There is no simple way to define the obviously related ideas of reputation, prestige, and influence. Reputation seems to be the broadest concept. In one sense, it ties to the idea of recognition—that is, the difference in the degree to which some people or institutions are more conspicuous, or more noticed, than others. But reputation is not just recognition. A person might be well known because of a well-publicized ad campaign, but it would not necessarily follow from this that he or she also has a reputation. What reputation adds to the concept of recognition is a glimpse of some part of one’s character. If someone has a reputation for honesty, that means that he or she is known for being honest. Conversely, if someone has a reputation for dishonesty, that means he or she is known for being dishonest. Reputation captures some part of an individual’s character and projects that part into the future. It is a way of reckoning some part of what

---

some individual has been as an index of what he or she will be. It follows that reputation may be negative or positive.\textsuperscript{10} And in this way, we capture the difference between reputation on the one hand, and prestige and respect, on the other. For one may have a good or bad reputation as a judge, but only if one has a good reputation does one also have prestige, or respect, as a judge. One may have prestige for reasons other than a good reputation—prestige may come from associations with prestigious institutions or people.

Further parsing of these concepts is unnecessary for present purposes. Suffice it to say that, in this context, prestige typically means “the amount of respect, regard, or esteem” a judge enjoys among other judges or interested publics.\textsuperscript{11} Since not all judges are equally prestigious, the concept has an inherently comparative and hierarchical element: some judges will inevitably be regarded as more prestigious, or as enjoying a higher reputation than others.\textsuperscript{12} Influence, in contrast, is “the extent to which the actions of one person have an effect on the views or behavior of others.”\textsuperscript{13} An influential judge is one whose opinions, or other work product, has affected the thinking or work product of the judges or other actors in the legal community.

In a simple world, we might imagine a simple linear relationship between legal reputation, prestige, and influence. A good judge would gain, by being a respected judge, prestige. Prestige and respect would build reputation, and once established, a good reputation would translate into influence. Influential judges would be those with the best reputation, and those with the best reputation would be those that were most influential. This simple world is not the one we inhabit. For one thing, the causal arrows between these concepts can get complicated. For example, a judge with low prestige or reputation (or none at all) can come to be influential because other judges follow her lead simply due to stare decisis or because that judge fortuitously rendered decisions that, for good or ill, have high precedential value. More important, shot through all these notions of judgment are conceptions of value, many of which are highly contested in our legal culture. To the extent that conceptions of value differ among individuals or among groups or classes of individuals, there are differences in views about good judges, or respected judges, or courts that enjoy good reputations. In short, an element of subjectivity becomes inherent in any discussion of reputation, prestige, and even influence.

\textsuperscript{10} Posner, supra note 9, at 58.
\textsuperscript{11} Klein & Morrisroe, supra note 9, at 371-72.
\textsuperscript{12} Sandefur, supra note 9, at 383-84.
\textsuperscript{13} Klein & Morrisroe, supra note 9, at 372.
There is, then, no simple or uncontentious way to describe the great judge or to mark out the most respected or influential courts. But this has not stopped people from talking about great judges or speaking of influential courts. Typically, such judges served on the Supreme Court, state high courts, or the lower federal courts, and they all produced work products—published written opinions—that are more or less permanent, accessible at relatively low cost, and in effect are required (or strongly recommended) reading due to stare decisis.\(^{14}\) In the past, scholars have examined attributes of what can be called the judicial craft. It can include the literary quality of opinions, the contribution the decisions make to the development of legal rules and principles, creativity, and skillful legal analysis, among other things.\(^{15}\) Still other factors are said to affect judicial reputation, some of which are largely or completely outside the control of the judge, such as longevity, pre-court or post-court careers, the proximity in time of the judge’s career to the person gauging the judge’s reputation, personal integrity, and ability to persuade other judges (in a multimember court).\(^{16}\)

Of course, many, perhaps most, of these factors would be considered by many people today as difficult to accurately or coherently measure, and hence subjective, or merely in the eye of the beholder.\(^{17}\) In a similar vein, the correctness of decisions by the judges, in the eyes of the evaluator, might be said to determine reputation, in whole or in part.\(^{18}\) One way around the skepticism about measuring reputation and related concepts is to look for correlative objective measures. It is no surprise, then, that citation analysis has been embraced in various contexts as the desired objective measure. In the academic community, citations can be used to measure the influence of prior published work. Professional norms require that later work build upon and, more importantly for us, refer to the earlier work. The same holds true for legal academic work and court opinions. Ci-

---

14. By contrast, compare the environment for a smart and thoughtful trial judge. Not all of her decisions may be rendered by written and accessible opinions; those that are not published in official reporters bind no one and thus suffer in influence. None of this is to say that the judge cannot enjoy a high reputation; but it will likely be restricted geographically to the lawyers that practice before her, and perhaps to nearby judges.


16. Ross, supra note 15, at 411-42 (discussing these and other factors).


tations, then, have come to be used as a convenient proxy for reputation, prestige, and influence.19 "Presumably, the best opinions will be cited more often than others."20

Citation analysis would seem to best serve as the measure of influence since frequent citation would, of itself, not necessarily translate to prestige or reputation. Yet even in that arena, the limits of citation analysis are well documented. Some of the possible drawbacks include: most studies do not distinguish between favorable or critical citations; many citations may reflect little more than the longevity of a judge’s career or the age of an opinion; a citation may be used to bolster an already-reached result; many citations are buried in string cites or are the result of self-citation; certain “superprecedents” on long-established and well-accepted principles may cease to be cited at all; and given that legal opinions are a public good, there are no costs associated with citing more or fewer cases on a particular point, so small differences in the perceived quality of opinions may lead to disproportionate disparities in the citation of cases.21

For many of the same reasons, citation analysis is at best an imperfect measure of the reputation of the authority being cited. Simply being cited a great deal does not mean a particular decision (or the judge that authored it) enjoys a high reputation, and the reverse is true as well. On the one hand, a particular decision may be the subject of frequent citation largely because it deals with an oft-litigated topic, not because it is a particularly admirable discussion of the topic.22 On the other hand, as just observed, a particular case, due to its landmark quality, may come to be so embedded in legal thought that later citation is thought to be less necessary or not necessary at all.23 All of which is to say that the quantity of citations does not automatically translate into a barometer of the quality of the work being cited—though it is not unrelated either. It is difficult to believe

22. A study of frequently cited U.S. courts of appeals decisions illustrates this point. See Robert Schriek, Most-Cited U.S. Courts of Appeals Cases from 1932 Until the Late 1980s, 83 LAW LIBR. J. 317, 322-23 (1991). Of the nineteen most-cited cases decided after 1932, the top three were decisions that, respectively, clarified the standard for deciding a directed verdict motion in a civil case; set out criteria for awarding attorney’s fees in civil rights actions; and held that decisions of the former Fifth Circuit were precedent for the then-new Eleventh Circuit. Id. (discussing Boeing Co. v. Shipman, 411 F.2d 365 (5th Cir. 1969); Johnson v. Georgia Highway Express, Inc., 488 F.2d 714 (5th Cir. 1974); and Bonner v. City of Prichard, 661 F.2d 1206 (11th Cir. 1981)).
that a particular decision, thought brilliant by some or many, would not undergo significant citation later.

Another reason for caution when using citation analysis, or any other measure, to study judicial reputation is the scope of the audience. Many studies are not entirely clear on whose opinion of reputation is being directly, or indirectly, studied. Within the confines of citation analysis, when examining the decisions of an intermediate appellate court, such as the U.S. courts of appeals, do we look to citations by the Supreme Court? by courts of appeals (intra- or intercircuit) themselves? by federal district courts? by legal academics? by nonlegal policymakers? All of these audiences have their own differing perceptions of the reputations of the courts of appeals.24 Given the relatively low public profile that the courts of appeals as a whole, and the circuit judges in particular, have compared to, say, the Supreme Court,25 probably the judges, practitioners, and academics who closely follow court of appeals decisionmaking would be the most fruitful audience to survey.

A final point is to consider whether it is meaningful to speak of the reputation of a multimember court as a whole, as opposed to the reputations of the judges who constitute that court. With rare exceptions,26 the studies of legal reputation deal with a particular judge or a particular case, not with an entire court as such. One might conclude that there is no useful distinction between the two concepts. Speaking again of the U.S. courts of appeals, “one could argue that a circuit’s influence is nothing more than the average of the influence of its judges.”27 Even if the circuit can be said to possess a reputation separate from the judges, “it would be difficult to disentangle it from the influence of the individual judges in the circuit.”28 If one circuit follows a case from another, how do we know if the invocation is to

27. Landes, Lessig & Solimine, supra note 21, app. at 327.
28. Id.
acknowledge the author of the opinion or the reputation of the circuit from which it came, or both? 29

Yet, there are several reasons to suggest that there are meaningful distinctions between the reputations of judges and the courts they compose. First, it is common, I think, in legal discourse to refer to the “court’s opinion” or the “court’s decision” more often than the decision of a particular judge. This may be for convenience, or it may reflect a lack of knowledge of the author of the opinion; either way it suggests that the court, as a collective entity, has a reputational existence apart from that of its members (with the reverse being true as well).

Second, there are institutional constraints in place that, at least in part, suggest that a circuit is an entity separate from the judges, and indeed from other circuits. Start with the fact that (with the exception of the Federal Circuit) the circuits have nonoverlapping geographic boundaries. All of the judges are paid equally and seniority plays little role, other than to determine who will serve as chief judge. 30 With regard to decisionmaking, most of the work is done via three-judge panels, subject to override by the circuit sitting en banc. 31 Those panels are typically constituted by random selection, 32 and the membership of the panels are often not revealed to counsel until shortly before oral argument. 33 With a few exceptions, three-judge panels must follow circuit precedent. 34 While non-circuit judges (that is, visiting judges from other circuits or district judges sitting by designation) are members of a significant number of three-judge panels, 35 their decisions still carry the authority of the circuit. Finally, most courts of appeals decisions are signed by the presumed lead author; the exceptions are unsigned short orders (often not officially

29. One way to partially deal with this problem is through the use of “invocation” studies to determine those occasions where a court gratuitously refers to the name of the judge who authors the opinion. See Part III, infra, for a discussion of these studies.


31. See id. § 46(c) (providing for en banc procedure). For discussion of en banc process, see Tracey E. George & Michael E. Solimine, Supreme Court Monitoring of the United States Courts of Appeals En Banc, 9 SUP. CT. ECON. REV. 171, 176-80 (2001).


33. E.g., 5TH CIR. I.O.P. 34 (one week before oral argument); 6TH CIR. I.O.P. 34(c)(2) (two weeks before oral argument); 11TH CIR. R. 34-4, I.O.P. 7 (one week before oral argument or earlier as determined by the court). See generally Brown & Lee, supra note 32, at 1075-78 tbl. 1 (summarizing notification practices from all of the circuits).


published) or per curiam opinions. The long tradition has been for decisions to be reached and opinions to be drafted in a collegial manner, that is, they are not simply the work product of one judge.

Third, it seems unlikely that circuits qua circuits will engage in strategic behavior when it comes to burnishing their reputation. Even if individual judges might be inclined to do this, to imagine circuits inflating citations to other circuits, expecting increased citation in return, imagines extensive collective action. The high cost of such collective action that would be required for two or more circuits to collude makes it extremely unlikely that circuit citation practice is in any meaningful sense exaggerated.

III. MEASURING THE REPUTATIONS OF THE UNITED STATES COURTS OF APPEALS

In this Part, I consider various sources that, directly or indirectly, can be said to gauge the historical and contemporary reputations of the federal courts of appeals. In Part IV, I will offer explanations for some of the patterns that emerge. I focus on the courts of appeals, since there has been relatively little attention given them, as collective entities, in the emerging literature on judicial prestige and influence. And I focus on reputation, because it too has not been the subject of extensive inquiry. There has been a good bit of citation analysis, but as already mentioned, it is a tool that best measures influence. Which is again not to say that the concepts are unrelated or that statistical or other measures of influence cannot be brought to bear when thinking about prestige or reputation. They do, as will soon be demonstrated.

The history and current institutional status of the courts of appeals is familiar, and only a brief overview is necessary here. The courts of appeals, as we know them, began business in 1891. Prior to that, there were no intermediate appeals courts in the modern sense in the federal system. The 1789 Judiciary Act had established district and circuit courts. The latter had original jurisdiction over some matters and appellate authority over certain types of civil and other cases. The circuit courts, however, did not have their own judges. They were staffed

---


38. Choi & Gulati, supra note 20, at 308-09 (arguing that strategic behavior by judges in this regard is unlikely given the existence of frequently rotating three-judge panels and norms of cooperation).

39. Landes, Lessig & Solimine, supra note 21, app. at 327.
by district judges and Supreme Court Justices sitting on circuit.\textsuperscript{40} There initially were three geographically distinct circuits,\textsuperscript{41} but other circuits were added as the nation’s territory expanded, eventually reaching nine.\textsuperscript{42} Judges who specifically sat on the circuits were eventually authorized as well,\textsuperscript{43} but various problems with the lack of a formal intermediate appellate court\textsuperscript{44} led to their creation by the Evarts Act in 1891.\textsuperscript{45} The new circuit courts of appeals had their own judges and were geographically distributed in a manner very similar to today. Formal changes since then have been few; the present Tenth Circuit was carved out of the Eighth in 1926, and the present Eleventh Circuit was carved out of the Fifth in 1980. In 1982, Congress created the Federal Circuit, which sits in Washington, D.C., but has national appellate jurisdiction over patent and various other specialized cases.\textsuperscript{46}

Today, many formal aspects of the thirteen circuits are uniform, such as pay and the disposition of most cases by three-judge panels. But some things are different. The number of states and their populations differ,\textsuperscript{47} as does the number of authorized active judges, which currently ranges from six (First Circuit) to twenty-eight (Ninth Circuit).\textsuperscript{48} No doubt, many other things could be added to the

\begin{tabular}{|c|c|}
\hline
CIRCUITS & NUMBER OF JUDGES \\
\hline
D.C. & 12 \\
1st & 6 \\
2d & 13 \\
3d & 14 \\
4th & 15 \\
5th & 17 \\
6th & 16 \\
7th & 11 \\
8th & 11 \\
9th & 28 \\
10th & 12 \\
11th & 12 \\
Federal & 12 \\
\hline
\end{tabular}

40. \textsc{Richard H. Fallon, Jr. et al., Hart & Wechsler’s The Federal Courts and the Federal System} 29 (5th ed. 2003) [hereinafter \textsc{Hart & Wechsler}].
41. \textit{Id.}
42. \textit{Id.} at 35.
43. \textit{Id.} at 36.
44. \textit{Id.} at 36-37.
45. Act of Mar. 3, 1891, ch. 517, 26 Stat. 826. The old circuit courts remained in existence, but their appellate jurisdiction was abolished; the circuit courts themselves were entirely abolished in 1911. \textsc{Hart & Wechsler, supra} note 40, at 37 n.65.
47. As of 1998, the population within each circuit ranged from 528,964 (D.C.) and 13,337,709 (1st) to 30,236,545 (6th) and 51,453,880 (9th). Five of the circuits cover three states; the Ninth Circuit covers eleven states. \textsc{Commission Report, supra} note 46, at 27 tbl.2-9.
48. The current breakdown is as follows:
list.49 One thing that has differed, at least starting in the 1920s,50 has been the reputation of the various circuits. From that time, up to the present, there is considerable, if uneven, evidence, culled from the popular press and various other sources, in which one or two circuits have had excellent reputations, above the rest, in various quarters of the legal community.

Over much of the twentieth century, the conventional wisdom has ranked the Second and District of Columbia Circuits highest with regard to prestige and influence. For example, in 1974 Washington journalist Joseph C. Goulden, in his entertaining survey of federal judges, labeled the D.C. Circuit as “the second most important court in the United States.”51 According to Goulden, this was due to “the court’s jurisdiction over appeals of decisions of a host of key federal regulatory agencies” and the fact that few such cases are reviewed by the Supreme Court.52 Similar comments in the modern press are not difficult to find.53 During the New Deal, President Roosevelt presciently noted that the D.C. Circuit had “taken on a wholly new importance in the last few years [and] is now easily the second most important Federal Court in the country.”54 Many judges on that court have been drawn from a national pool,55 and over the years the D.C. Cir-


50. Though I make no claim to have thoroughly canvassed the possible historical sources, I have found relatively few discussions in the sources consulted and cited in this Essay of the reputations of the courts of appeals from 1891 to the 1920s or, for that matter, for the circuit courts prior to 1891. I tend to think there were differing reputations, and that at least to some degree they influenced the modern reputations discussed in this Essay. For example, the distinctive reputational flavor of the modern D.C. Circuit might be traced, in part, to the unique federal trial and appellate courts that Congress has created for the District of Columbia since 1789. See CHRISTOPHER P. BANKS, JUDICIAL POLITICS IN THE D.C. CIRCUIT COURT 7-10 (1999) (tracing the complicated historical roots of the modern D.C. Circuit). On the other hand, perhaps the phenomenon of the appellate court reputation is closely linked to the appointment of Learned Hand to the Second Circuit in 1924. See infra note 60.


52. Id.; see also BANKS, supra note 50, at 2-3.


cuit has been a source of jurists seriously considered by Presidents for appointment to the Supreme Court, with several actual appointments.\(^{56}\)

Similarly, the respect historically given the Second Circuit is often attributed to the brilliance of the judges who have served on the court, as well as to the commercial and other important issues decided by that court.\(^{57}\) These accounts were echoed by federal judges\(^{58}\) and academic commentators.\(^{59}\) Even more so than the D.C. Circuit, I think, the Second Circuit’s reputation is closely aligned with that of the well-known judges that served on it—starting with Learned Hand, who served from 1924 to 1961,\(^{60}\) but also including Charles Clark, Jerome Frank, Augustus Hand, and Henry Friendly, among others.\(^{61}\)

---

56. Banks, supra note 50, at 4 (noting that one-third of the current Court formerly sat on the D.C. Circuit (Justices Ruth Bader Ginsburg, Scalia, and Thomas) and that other judges have either been nominated (Bork) or placed on short lists (Douglas Ginsburg) for consideration). It is also worth mentioning that Chief Justices Fred Vinson and Warren Burger also served on the D.C. Circuit. Id. at 125.

57. E.g., Margaret A. Jacobs, Court Finds Limits on Insurance Held by Asbestos Firms, WALL ST. J., May 18, 1994, at B7; The Talk of the Town: George’s Choice, NEW YORKER, Jan. 18, 1993, at 31, 31 (“T[he Second Circuit is often called the second most important court in the nation.”). Indeed, one measure of influence or reputation could be to determine the comparative coverage or mention of the circuits in the popular press, cf. Lee Epstein & Jeffrey A. Segal, Measuring Issue Salience, 44 AM. J. POL. SCI. 66 (2000) (measuring coverage of Supreme Court decisions in The New York Times), or in the legal press (for example, determining how often the circuits are mentioned in The National Law Journal or have decisions summarized in United States Law Week).


59. Id. at xix (referring to the Second Circuit as “the nation’s leading commercial court”); Marvin Schick, Learned Hand’s Court 5 (1970) (referring to the Second Circuit “as one of the top appellate courts in the history of the country”).

60. It is difficult to overstate the prestige of Judge Hand. He was the subject of a magisterial biography eleven years ago, Gerald Gunther, Learned Hand: The Man and the Judge (1994), which spawned further discussion of judicial biography in general and of Hand in particular. See Symposium, National Conference on Judicial Biography, 70 N.Y.U. L. REV. 485 (1995); Posner, supra note 15; Przybyszewski, supra note 17. Virtually every modern discussion of judge and court reputation makes some mention of Judge Hand, over four decades after his death.


Frank was a prominent New Deal lawyer and a famed legal realist. See Robert Jerome Glennon, The Role of a Circuit Judge in Shaping Constitutional Law: Jerome Frank’s Influence on Supreme Court, 1978 ARIZ. ST. L.J. 523, 524-25.

Augustus Hand was Learned’s first cousin, and he was highly thought of in his own right. Augustus sat on the Second Circuit from 1927 to 1954. For further discussion of the tenure of the Hand cousins on that circuit, see Michael E. Solimine, Nepotism in the Federal Judiciary, 71 U. CIN. L. REV. 563, 579-80 (2002).
Yet until very recently, there had been little effort to empirically inform these assumptions, particularly, though not only, through the comparison of one circuit (or particular judge) to another. In one study, Marvin Schick, focusing on the decades Learned Hand served on the Second Circuit, examined how often the circuits were reviewed by the Supreme Court. He found that between 1941 and 1951 the Second Circuit was reviewed the most but reversed the second least; both findings were, according to him, indicia of the circuit’s prestige and influence. In another study, J. Woodford Howard examined citation practices in three circuit courts (D.C., Second, and Fifth) from 1965 to 1967. He found, among other things, that the Ninth Circuit was cited most often by the other three circuits, while the rest of the circuits (including the Second) were cited at about the same rate by each other. The D.C. Circuit came in last. A more recent study examined the most-cited courts of appeals cases from mid-1932 to mid-1988. Of the nineteen most-cited cases, six were from...
the Second Circuit (including two authored by Learned Hand), and six were from the D.C. Circuit.67

These studies are interesting and important and, as we shall see in a moment, served as useful models for more recent work. And they largely support the conventional wisdom of the prominence of the Second and D.C. Circuits. Nonetheless, they are not definitive given their age and the limited coverage of judges, courts, or decisions canvassed. It is to the more recent studies that I now turn.

First consider the view of attorneys. One helpful source is the *Almanac of the Federal Judiciary*, which profiles each circuit and quotes attorneys who litigate before those courts.68 Not unexpectedly, attorneys in the Second and D.C. Circuits gave these circuits the highest compliments.69 But attorneys in almost all of the other circuits also gave their own courts the highest compliments, sometimes comparing themselves favorably to the Second and D.C. Circuits.70 The weight given these remarks must be discounted by the small sample size, their anecdotal and probably parochial nature, and the

---


68. The *Almanac* mentions that it surveys a “cross-section” of attorneys, but it makes no claims of having made a scientific survey of lawyers, or of having polled them with the same questions. See 1 ALMANAC OF THE FEDERAL JUDICIARY, at Intro. 1 (2005).

69. For comments on the Second Circuit, see 2 id. at 2d Cir. 1 (“definitely one of the better circuits”; “[t]he Second Circuit is top notch, the best in the United States”; “the circuit has been a preeminent leader on business issues”). For comments on the D.C. Circuit, see 2 id. at D.C. Cir. 1 (“one of the best [court of appeals] in the country”; “[a]lmost all of these judges are way above the average judge”; “[t]hese judges have all developed an expertise in administrative and regulatory law”; “uniformly impressive”).

70. See, e.g., 2 id. at 1st Cir. 1 (“very scholarly”; “outstanding intellectual body”); 2 id. at 5th Cir. 1 (“[O]n the whole, the Fifth Circuit is on equal level with New York and D.C. and in the top 10-percent or so.”); 2 id. at 7th Cir. 1 (“excellent court,” “one of the best courts in the country,” particularly emphasizing presence of Judges Posner and Easterbrook); 2 id. at Federal Cir. 1 (“The court’s legal ability is very good.”). But see 2 id. at 6th Cir. 1 (“It’s a solid circuit, but not scholarly.”); 2 id. at 10th Cir. 1 (“But if the Tenth Circuit still lags behind others, such as the Seventh Circuit, we’re not behind by a big margin.”).
fact that many of the attorneys surveyed did not have extensive appellate experience in other circuits, thus making comparisons difficult.

Next, consider David Klein’s recent study of the courts of appeals, in which he reports his interviews of twenty-four appellate judges from six circuits.\textsuperscript{71} The respondents were asked whether, in deciding cases, they gave particular weight to the decisions of certain judges or other courts.\textsuperscript{72} The bottom line answer is, for most judges in most cases, not much. Most of them gave weight to what positions other circuits had taken on a particular issue, how many circuits had reached the issue, and whether there was a circuit split. The identity of the circuits was typically less important.\textsuperscript{73} Klein reports the following “typical” reactions to questions on whether circuits have discernable reputations:

I really don’t think so. When I was young, there was the Second Circuit, with Hand, Frank, and Swan. All the circuits are bigger now with their membership constantly changing. When you get circuits that big, it’s not likely that you can have star courts.\textsuperscript{74}

That’s a will-o’-the-wisp I wouldn’t trust as far as I could throw it. When you back away and look, it’s hard to say.\textsuperscript{75}

No. That’s a dichotomy between the public and judges. Most federal judges would feel no one circuit is better than the others. The reasons: One, composition changes rapidly. There may be good judges on it now, but bad judges may join, the good may leave. Two, the difference in the size of the circuits. Some are so big they can’t have a single reputation.\textsuperscript{76}

In contrast, some of the judges did react negatively to the reputation of the Ninth Circuit:

The Ninth Circuit, of course, though it has great individual judges, is so large and its jurisprudence is so diversified over a tremendous area that it doesn’t have the same jurisprudential integrity I think we have.\textsuperscript{77}

\textsuperscript{71} David E. Klein, Making Law in the United States Courts of Appeals 18–19 (2002) (outlining methodology of interviews). Ten judges were from the Sixth Circuit, five from the Third Circuit, four from the Seventh Circuit, two each from the First and Eleventh Circuits, and one from the Fourth Circuit. \textit{Id.} at 18, 19 n.10.

\textsuperscript{72} Klein asked the judges if “there are any circuits today which have a reputation for general excellence or which in your view merit such a reputation,” and if they would be “more inclined to adopt rules from certain courts or judges than others.” \textit{Id.} app. B at 168-69.

\textsuperscript{73} See \textit{id.} at 57, 89-92.

\textsuperscript{74} \textit{Id.} at 92.

\textsuperscript{75} \textit{Id.}

\textsuperscript{76} \textit{Id.}

\textsuperscript{77} \textit{Id.}
I’ve seen a tendency on the part of judges to recoil from, reject anything from the Ninth Circuit, because they’re way out there, do a lot of experimenting with the law. There are a lot of knee-jerk reactions by other judges; they tend to discount any precedent from the Ninth—consider it too liberal, activist.78

In sum, according to Klein, “Judges typically do not think of whole circuits in evaluative terms and so do not weight precedents according to the circuit they come from.”79

Klein’s respondents had some different views on the authors of decisions from other circuits, as “most felt that the name on the opinion did affect their decision making at times.”80 As reported by Klein, some of their reactions were as follows:

Once in a great while I feel an initial kick because it’s from a great judge—say Friendly or Wisdom. But usually it’s the opinion itself, if it’s a thoughtful opinion.81

I guess it might matter. I wouldn’t distinguish by circuit. There are certain judges I know and have a lot of respect for. If I find they said something, I might give it a little more weight. Maybe “weight” is not the right word. I think through the case myself, but if a judge I respect agreed exactly with my position, I’d feel more satisfied, while if that judge were diametrically opposed I would pause.82

Sure, the better the judge, the more seriously you take them. Some people you know personally, others just through opinions, but you form a sense of how good they are through their work.83

I think so. If it’s a judge I know or who is reputed for his scholarship or legal acumen, I will probably give greater deference than if the judge is unknown or has a lesser reputation.84

Oh, yeah. There are some I think I’m more simpatico with. Also, I certainly take note of ones from Posner. I’m impressed by Kearse, Oakes, and some others on the Second Circuit. This is factored in almost unconsciously. Judge Winter is just too conservative. He’s supposed to be a fine judge, but I’m not very impressed.85

Even if there isn’t a definitive body of law in other circuits but there’s an opinion by a judge, even a dissent—I respect some judges more than others. . . . For example, Noonan on ethics, Pos-

78. Id. at 92-93.
79. Id. at 93.
80. Id. at 94.
81. Id.
82. Id.
83. Id.
84. Id.
85. Id. at 95.
ner on economics. I would want to pick his brain. On the Fourteenth Amendment, Bork.\footnote{See Schick, supra note 59, at 336.}

The number of cases reviewed from a circuit might be considered to be some indication of the importance of the cases decided by that circuit. Likewise, the higher the rate of affirmance by the Supreme Court in those cases, the more deference the Court is giving to that circuit, which might be some indication of the circuit’s prestige. For short periods of time, these factors might be meaningful. Much ink has been spilled lately, for example, on the disproportionately large number of cases from the Ninth Circuit reviewed, and often reversed by, the Supreme Court.\footnote{According to data compiled by Lee Epstein and her colleagues, the number of decisions reviewed on the merits from each circuit, from the 1946 through the 2001 Terms, ranged from 127 of the Eleventh Circuit (only operating from 1981) to 759 of the Ninth Circuit. The rates of affirmance ranged from a low of 30.9% (D.C. Circuit) to a high of 47.8% (Second Circuit). See Epstein et al., supra note 31, at 174-75 (providing a brief review of scholarly literature on the certiorari process). It is not obvious how these factors, other than the first and perhaps the last, directly speak to a circuit’s reputation. Indeed, the denial of certiorari can be regarded in some circumstances as the Court being deferential to the court below, thus reflecting well on the court or the author of the opinion. Glennon, supra note 61, at 537 (giving example of Justice Harlan telling Learned Hand that a denial of certiorari in a case the latter authored really meant “Judgment Affirmed”). That said, perhaps review and affirmance as indicia of reputation carried more meaning when the Court was deciding more cases, and the courts of appeals fewer, than in recent decades.}

Yet it is not clear that it has somehow lessened the reputation of that circuit. Moreover, over time, the rates of review and reversal of circuit opinions are quite similar,\footnote{Even if there were meaningful differences between the circuits in this regard, it is doubtful that Supreme Court review and affirmance would be a good surrogate for reputation. The Court, no doubt, considers a host of factors when deciding whether to grant certiorari in a case. These factors can include the perceived importance of the issue presented in the case; the presence or absence of a conflict with other circuits; how far the issue may have percolated among the circuits; the need for uniformity on this issue of federal law; and the correctness of the decision below. See George & Solimine, supra note 31, at 174-75 (providing a brief review of scholarly literature on the certiorari process). It is not obvious how these factors, other than the first and perhaps the last, directly speak to a circuit’s reputation. Indeed, the denial of certiorari can be regarded in some circumstances as the Court being deferential to the court below, thus reflecting well on the court or the author of the opinion. Glennon, supra note 61, at 354 (giving example of Justice Harlan telling Learned Hand that a denial of certiorari in a case the latter authored really meant “Judgment Affirmed”). That said, perhaps review and affirmance as indicia of reputation carried more meaning when the Court was deciding more cases, and the courts of appeals fewer, than in recent decades.}

This is not to say that other aspects of the Supreme Court-circuit court relationship may not yield insights regarding circuit reputa-
tion. As one example, Court opinions sometimes specifically refer to the author of the opinion being reviewed (or of a dissenting opinion, or the author of some other, related case), even when citing conventions do not require such a reference. This presumably reflects the greater authoritative weight that the Court wishes to give to the lower court opinion being referenced. Thus, in recent decisions, one finds specific reference in Court opinions to Judges Richard Posner, Frank Easterbrook, Robert Bork, and Harry Edwards (among living jurists), and to Judges Learned Hand and Henry Friendly (among jurists who have passed away).91 Even here, though, the references are usually to judges, not to a three-judge panel as such.92

As another example, almost all of the clerks for Supreme Court Justices are recent law school graduates who have clerked for a court of appeals judge. Given the high prestige afforded those positions, the feeder judges (or circuits) for whom the clerks previously worked might also be afforded a high reputation. During the 1975 through 2003 Terms, the circuits supplying the most clerks were the D.C. (36.4%), Ninth (15.1%), and Second (14.9%).93 No other circuit was above ten percent.94

Finally, several recent examples of citation analysis can shed light on the reputation of the circuits. The first is by William Landes, Lawrence Lessig, and myself.95 We examined the citations, by other


93. E-mail from Lawrence Baum to Michael E. Solimine (July 19, 2004) (on file with author).

94. See id. The circuit breakdown is as follows: D.C. (36.4%); Ninth (15.1%); Second (14.9%); Fourth (9.9%); First (5.9%); Fifth (5.3%); Seventh (5.0%); Third (3.3%); Eleventh (1.5%); Tenth (1.2%); Eighth (1.0%); Sixth (0.5%). Id. The information supplied by Professor Baum is a follow-up to his article, Corey Ditslear & Lawrence Baum, Selection of Law Clerks and Polarization in the U.S. Supreme Court, 63 J. Pol. 869 (2001).

95. Landes, Lessig & Solimine, supra note 21.
courts of appeals, of published opinions authored by 205 appellate judges, with at least six years of service, sitting as of 1992. The database included all published opinions by those judges; the earliest was rendered in 1955, the latest in 1995. Most pertinent for the present discussion, we ranked all of the judges by the number of citations to their opinions outside of their circuit. The top ten consisted solely of judges on the Seventh, First, and Second Circuits, with Judges Posner, Bruce Selya, and Easterbrook heading the list. We also aggregated the judge-specific data by circuit, based on opinions published from 1982 to 1995. Once again, the most pertinent data from this part of the study are citations by other circuits, adjusted for the number of opinions generated by the circuit being cited. On that score, the Third Circuit, followed by the Second, D.C., and Seventh Circuits had the greatest influence.

A second citation analysis study, by David Klein and Darby Morrisroe, examined how often courts of appeals judges were cited or mentioned by name in courts of appeals opinions. This is parallel to the earlier-described invocation studies, which examine how often circuit judges are mentioned in Supreme Court opinions. The authors examined references to a sample of 139 circuit judges sitting in the late 1980s and early 1990s. The judges with the most references were William Wilkins (Fourth Circuit), Stephen Breyer (then of the First Circuit), Posner, and Easterbrook. The frequent references to Judge Wilkins were “attributable largely to his service as chair of the United States Sentencing Commission.”

Stephen Choi and Mitu Gulati provide a third use of citation analysis to gauge circuit reputation. In the course of developing and applying criteria that seek to objectively measure the quality of a circuit judge’s opinions, Choi and Gulati studied (among other things) the citation of opinions published from 1998 to 2000 by circuit

96. Id. at 276-79 (describing methodology).
97. Id. at 288-92 tbl.2A. The top ten were as follows: Posner (Seventh Circuit); Selya (First Circuit); Easterbrook (Seventh Circuit); Coffin (First Circuit); Campbell (First Circuit); Cudahy (Seventh Circuit); Newman (Second Circuit); Bownes (First Circuit); Flaum (Seventh Circuit); Oakes (Second Circuit).
98. To compare just the circuits, we started in 1982, since that was the year the 11th Circuit began operations and publishing opinions. Id. app. at 328 n.72.
99. Id. app. at 332.
100. Klein & Morrisroe, supra note 9.
101. For further discussion of methodology, see id. at 377-80. The 139 judges were a sample of all active or senior circuit judges at the time, limited to those who wrote majority opinions in “287 circuit court cases involving issues not previously settled by the Supreme Court.” Id. at 377.
102. Id. at 381 tbl.2.
103. Id. at 382.
judges sitting as of June 2003. The top five judges, ranked with regard to citations to their opinions outside of their respective circuits, were Posner and Easterbrook (Seventh Circuit); Sandra Lynch and Selya (First Circuit); and Paul Kelly (Tenth Circuit). By partially aggregating the data by circuit and controlling for number of opinions, some of the circuits (such as the First and Third) have a high percentage of their judges with relatively high citation counts. In contrast, the D.C. Circuit was at the bottom of the list.

What conclusions can be drawn from these various studies? Any conclusions should be drawn with care. Reputation is a difficult subject to objectively study. Couple that with the snapshot quality of most of the studies; they usually cover a relatively short period of time or only samples of the judges who constitute a circuit. That all said, at one point there was support for the conventional wisdom that the reputation of the Second and D.C. Circuits towered over the rest. However, in the past several decades, it seems, the reputations of those circuits have fallen to a degree, or perhaps one can say that the reputations of the other circuits have risen (or both). Put another way, there seems to be a homogenization of reputations among the circuits. There are exceptions to the point. One is the First Circuit, which perhaps quietly has had a high reputation in some quarters, as reflected to a degree in the good showing of judges from that circuit in my study with Landes and Lessig. A more notable exception is the Seventh Circuit, whose reputational stock has skyrocketed in the past two decades, related in no small way to the prestige afforded Judges Posner and Easterbrook, who joined that court in the 1980s.

IV. THE RISE AND FALL OF REPUTATIONS OF THE CIRCUITS

The prior Part demonstrates that the reputations of the U.S. courts of appeals are not static. In my view, the evidence shows that the prior high prestige of the Second and D.C. Circuits has, to some degree, lessened in the latter decades of the twentieth century. At the same time, the reputation of the Seventh Circuit has risen, and to varying degrees, the reputations of all of the circuits have risen. I draw these conclusions with caution; many of the studies previously discussed only focused on one court’s influence on another, not on

105. Id. at 40-41 (generally describing methodology).
106. Id. at 50 tbl.4. For one of their measures of opinion quality, the authors limited the study to the top twenty opinions of each judge to restrict the bias attributable to each circuit publishing differing percentages of opinions. Id. at 52-54. When controlling for number of opinions in this way, the order was Lynch, Easterbrook, Kelly, Posner, and Selya. Id. at 53 tbl.5.
107. Id. at 79 tbl.16.
108. Id.
reputation as such. None of the studies are longitudinal in nature, and all draw from, or are directed at, different audiences. That said, the general contours of circuit reputation possess enough clarity to be worthy of study and explanation. An appropriate starting point to sketch out reasons for reputational differences is to focus on attributes of the court whose reputation is being studied and on the relations between those courts and other audiences—other courts and judges, lawyers, and other observers in the legal community.110

First, consider the rise of the Seventh Circuit’s reputation. That ascension can be traced almost directly to the appointments by President Reagan of Richard Posner and Frank Easterbrook to that circuit in 1981 and 1985, respectively. Prior to that time, it is worth saying, “the Seventh Circuit was long considered an unremarkable circuit in terms of its work product.”111 But after those appointments, the entire circuit began producing more publishable opinions,112 a phenomenon that persists to the present day.113 The stories of Posner and Easterbrook are, of course, familiar ones. At the time of their appointments, both were well-known academic scholars at the high-profile University of Chicago Law School, particularly associated with the law and economics movement. They remain senior lecturers at Chicago and have produced a stream of scholarly work since then.114 The circuit’s reputation has been further burnished by the appointment of two other law school professors (Kenneth Ripple and Diane Wood).115 While the Seventh Circuit’s meteoric rise in reputa-

---

110. For a helpful discussion of this point, see Gregory A. Caldeira, The Transmission of Legal Precedent: A Study of State Supreme Courts, 79 AM. POL. SCI. REV. 178, 181 (1985). I am not claiming that these attributes and relational explanatory models are sharply distinct, for “most of the explanations have variants that come under both headings.” Id.

111. Gulati & Sanchez, supra note 4, at 1180 (drawing on conversations with judges and academics).

112. Id. at 1180-81.

113. In recent years the Seventh Circuit has published, on the average, nearly half of its appeals decided on the merits, a figure rivaled only by the First Circuit. The other circuits typically publish much lower percentages. See Jonathan Matthew Cohen, Inside Appellate Courts: The Impact of Court Organization on Judicial Decision Making in the United States Courts of Appeals 76-78 tbl.4 (2002) (summarizing data from the courts of appeals from 1990 to 2000). During the same period, the overall average of published opinions fell from 32% to 20%. See id. at 73-74.

114. A study of the legal scholars most cited in scholarship, using databases from 1956 to 1999, demonstrated that Posner and Easterbrook ranked first and twenty-first, respectively. Other courts of appeals judges in the top fifty were Guido Calabresi (tenth), Robert Bork (sixteenth), and Henry Friendly (twenty-seventh). Fred R. Shapiro, The Most-Cited Legal Scholars, 29 J. LEGAL STUD. 409, 424 tbl.6 (2000).

115. Ripple and Wood were on the faculties of Notre Dame and the University of Chicago law schools, respectively, and were Reagan and Clinton appointees, respectively. For a general study of the appointment of legal academics to the U.S. courts of appeals, see Tracey E. George, Court Fixing, 43 ARIZ. L. REV. 9 (2001). Perhaps the Seventh Circuit’s
tion is largely attributable to Posner and Easterbrook, it is not completely so. The norm for the entire circuit (or at least many of the judges thereof) has changed to one where more high-quality, publishable opinions are produced.

Mitu Gulati has suggested that the concept of circuit reputation can be profitably analyzed through the concept of circuit norms. Judges on multimember courts do not work in isolation. Over time, circuits appear to implicitly develop cultures that manifest themselves in various ways. These can include the ideological direction of decisions, whether oral argument is preferred, the swiftness of opinion writing and the length and style of opinions, how many are officially published, proclivity toward en banc review, and a host of other factors. No doubt, circuit norms influence judges who join the court and in turn are affected or generated by the judges themselves. Life-tenured federal judges surely are concerned at some level with maintaining or increasing their own reputation (or influence) and that of their circuit. Perhaps Posner and reputation would be even higher if University of Chicago law professor Antonin Scalia had been appointed. According to one source,

Scalia turned down the Reagan Administration’s first invitation to serve in the federal judiciary. Rather than accept the offer to sit on the U.S. Court of Appeals for the Seventh Circuit, Scalia waited patiently for an opening on the U.S. Court of Appeals for the District of Columbia, a more prestigious appellate court and one where his expertise in federal administrative law could be put to better use. JAMES F. SIMON, THE CENTER HOLDS: THE POWER STRUGGLE INSIDE THE REHNQUIST COURT 140 (1995) (emphasis added). The episode is similar to Dean Charles Clark resisting FDR’s entreaties to join the D.C. Circuit. Clark eventually accepted an invitation to join the Second Circuit. GOLDMAN , supra note 54, at 26-27. In contrast, Frank Easterbrook was more interested in appointment to a “regional” circuit, as opposed to the D.C. Circuit, in part because “[a]dministrative law is enjoyable, but a varied diet is better.” Howard Bashman, 20 Questions for Circuit Judge Frank H. Easterbrook of the U.S. Court of Appeals for the Seventh Circuit, HOW APPELLING (Aug. 2, 2004), at http://legalaffairs.org/howappealing/20q/2004_08_01_20q-appellateblog_archive.html.

117. For further discussion of circuit norms, see Mitu Gulati & C.M.A. McCauliff, On Not Making Law, LAW & CONTEMP. PROBS., Summer 1998, at 157, 161; and Gulati & Sanchez, supra note 4, at 1180-81, 1187. For a general discussion of the concept of norms in legal thought, see ERIC A. POSNER, LAW AND SOCIAL NORMS (2000).
118. RICHARD A. POSNER, OVERCOMING LAW 117-19 (1995) (discussing reputation, prestige, and other utility functions that federal appellate judges seek to maximize). Similarly, circuit judges are surely aware of their collective reputation, and it would seem that many would endeavor to maintain or improve upon it. See, e.g., Marci Alboher Nusbaum, A Court Known for Balance, Intellect, NAT’L L.J., Apr. 2, 2001, at B9 (“Judge Hand’s legacy is a powerful presence in the [Second] Circuit and one that the court has proudly tried to live up to.”). Perhaps some circuits, such as the First, Second, and D.C., have a greater sense of tradition and seek to maintain or inculcate it in various ways, more so than other circuits.

119. Gulati & McCauliff, supra note 117, at 200-01; Gulati & Sanchez, supra note 4, at 1180-81. There are other ways in which appellate judges can burnish their reputations. One would be contributions to scholarly, academic writing. Another might be for judges to hire excellent law clerks who, presumably, think highly of the judge and subsequently cultivate the judge’s image in academia or other circles. Cf. Barry Cushman, Clerking for
Easterbrook, explicitly or implicitly, made efforts to change norms on the Seventh Circuit.\footnote{120} Whether and to what extent such norms operate in other circuits now, or influenced judicial reputation in the past, are subjects worth exploring, but ones beyond the scope of this Essay. Nonetheless, we can highlight several factors which help explain the historical and, to some degree, continued high reputations of the Second and D.C. Circuits. Those courts share a number of characteristics which suggest why they, rather than, say, the Sixth or Tenth Circuits, are at the top in rankings of reputations. Most obviously, they sit in cities that are the major centers of attention in the United States. The D.C. Circuit is in the nation’s capital and, as noted earlier, has historically been staffed by judges from across the country. The Second Circuit sits in New York City, which without overstatement can still be regarded in many ways as the commercial, communications, and cultural capital of the country. Presidents have appointed the highest percentage of law professors to those circuits, which would tend to increase reputation (at least among academics).\footnote{121}

Another distinguishing, unique characteristic of the Second and D.C. Circuits is comparative regionalization.\footnote{122} In the other circuits, judgeships are allocated by state. In the Sixth Circuit, for example, Michigan has six judgeships, Ohio has four, Kentucky has three, and Scrooge, 70 U. CHI. L. REV. 721, 738-42 (2003) (book review). Cushman discusses clerks for Supreme Court Justices, but his point would seem to apply to lower federal court judges, as well. See Richard A. Posner, Judicial Behavior and Performance: An Economic Approach, 32 Fla. St. U. L. REV. 1259 (2005).

Judge Posner has expressed skepticism of the ideal of judges laboring to increase the prestige of other judges, which might translate to a skepticism of a judge making particular efforts to increase the prestige (or reputation) of the circuit:

Apart from opposing an increase in the number of judges or a dilution of the title “judge,” however, there is little an individual judge can do to enhance his judicial prestige. That prestige inheres in the whole judiciary. Free-rider problems make it unlikely that any one judge will exert himself strenuously to raise the prestige of all.

Posner, supra note 118, at 118. He is in a better position to judge than I, but my sense is that most of the relatively small number of lawyers who attain what is regarded as a prestigious position will take cooperative steps to enhance their collective reputation. As one example, consider the various efforts of the Judicial Conference of the United States and of some individual judges to oppose large-scale expansions of the number of life-tenured federal judges or to convert non-Article III (that is, non-life-tenured) judges to Article III status. Judith Resnik, Trial as Error, Jurisdiction as Injury: Transforming the Meaning of Article III, 113 HARV. L. REV. 924, 983-86 (2000). The opposition has been driven, at least in part, by the desire of federal judges that “the Article III judiciary must remain relatively small to retain the elite status that has traditionally lured first-rate lawyers to the federal bench.” Hart & Wechsler, supra note 40, at 385.

\footnote{120} Gulati & Sanchez, supra note 4, at 1180-81.
\footnote{121} George, supra note 115, at 44-45 (collecting data on appointments from FDR to Clinton). Law professors constituted 22.2% and 33.3% of the appointees to the Second and D.C. Circuits, respectively. The next highest circuit was the Tenth, at 14.3%. Id. at 45 fig.2.
\footnote{122} Thanks to Richard Posner for bringing the point to my attention.
Tennessee has three. Within each state, tradition or senatorial prerogatives may demand that some of those judges come from specific parts of the state. This alone narrows the field of selection. Contrast the D.C. Circuit, which is explicitly free of such constraints and indeed historically enjoys a norm of being filled by judges from across the nation. The Second Circuit has the majority (nine) of its judges from New York and has only a handful from the other states, Vermont (one) and Connecticut (three). And not coincidentally, the elite law schools in New York and Connecticut (for example, Yale) supply many of the Second Circuit jurists.

Both circuits have also historically had a disproportionate number of what are regarded as important cases on their dockets. For many decades in the past century, the Second Circuit considered more commercial, copyright, securities, tax, and antitrust cases, as compared to other circuits. The D.C. Circuit has long decided a disproportionate number of appeals from often complicated decisions from federal administrative agencies, including but not limited to environmental cases. These types of cases may be thought to be more

123. Judges are usually allocated to each state within a circuit based on the cases generated by each state. There must be at least one judge designated for each state within the circuit. See 28 U.S.C. § 44(c) (2000).


125. Similar points could be made about the First Circuit, many of whose judges have been from Massachusetts and which can draw on the venerable legal tradition of Boston and the graduates of Harvard Law School.

126. See Schick, supra note 59, at 187-88, 309; Lawrence Baum et al., The Evolution of Litigation in the Federal Courts of Appeals 1895-1975, 16 LAW & SOC'Y REV. 291, 298 tbl.3 (1981-82); James H. Carter, They Know It when They See It: Copyright and Aesthetics in the Second Circuit, 65 St. John's L. Rev. 773, 773 (1991) (“The Second Circuit is widely recognized as the nation’s most important copyright court.”); Erin B. Kaheny, Agenda Change in the U.S. Courts of Appeals, 1925-1988, 20 JUST. SYS. J. 275, 286 tbl.6-3, 287 tbl.6-4 (1999) (documenting that the Second Circuit historically has decided more intellectual property and antitrust cases than other circuits); Posner, supra note 15, at 513 (“Hand’s copyright opinions [are] generally considered Hand’s finest opinions taken as a group . . . .”).

127. Goulden, supra note 51, at 252-53; Howard, supra note 58, at 32; Stefanie A. Lindquist et al., Assessment of Caseload Burden in the U.S. Court of Appeals for the D.C. Circuit: Report to the Subcommittee on Judicial Statistics of the Committee on Judicial Resources of the Judicial Conference of the United States 9-10 (1999) (on file with author) (finding that in the years 1996 and 1997, appeals from federal administrative agencies made up more than forty percent of the D.C. Circuit’s docket; no other circuit exceeded ten percent); Sue Davis & Donald R. Songer, The Changing Role of the United States Courts of Appeals: The Flow of Litigation Revisited, 13 JUST. SYS. J. 323, 327 tbl.2 (1988-89) (reporting similar data from 1984); Richard J. Pierce, Jr., The Special Contributions of the D.C. Circuit to Administrative Law, 90 Geo. L.J. 779 (2002); see also Richard J. Pierce, Jr., The Relationship Between the District of Columbia Circuit and Its Critics, 67 Geo. Wash. L. Rev. 797, 797 (1999) (“I suspect that the D.C. Circuit is second only to the U.S. Supreme Court with respect to the volume of critical writing its opinions elicit. That high level of critical attention is to be expected independent of the quality of the court’s decisionmaking process. The court’s
intellectually interesting or demanding than the typical case that appears on a circuit’s docket. Some indirect support for these propositions is found in the study of Chicago lawyers by John Heinz and Edward Laumann. Seeking to examine the comparative reputation of different areas of practice, they asked a sample of lawyers and law professors to rate the prestige of fields of practice within the profession.128 Both groups listed securities, tax, antitrust, patent, banking, and public utilities as the leaders in prestige.129

Yet we should not make too much of comparative docket composition. Even if historically less noticed, other circuits have in effect developed their own specialities—for example, admiralty law in the Fifth Circuit and immigration and copyright law in the Ninth Circuit.130 And “[t]here is nothing particularly remarkable about the Seventh Circuit’s docket.”131 More generally, it appears that, over time, the docket composition of the circuits has become relatively more uniform, although there remain pockets of variation (with the D.C. Circuit’s administrative docket being the best example).132

128. HEINZ & LAUMANN, supra note 9, at 90-103 (describing the study). The professors in the sample were from Northwestern University. Id. at 101-02.

129. Id. at 91 tbl.4.1, 103 tbl.4.3. The Heinz and Laumann study was conducted in the late 1970s, but more recent surveys of the Chicago bar have yielded very similar results. Sandefur, supra note 9, at 386-87 tbl.1 (reporting the results of a survey conducted in 1995).

While patent practice ranks high in prestige, citation analysis of the influence of the Federal Circuit (which has exclusively heard appeals of patent cases since 1982) shows that it ranks last among the circuits. Landes, Lessig & Solimine, supra note 21, at 317 tbl.5. This result is not surprising given the highly specialized nature of that court’s work product and that it publishes the fewest cases of any circuit. Id. at 303. For evaluations of the court, see Rochelle Cooper Dreyfuss, The Federal Circuit: A Continuing Experiment in Specialization, 54 CASE W. RES. L. REV. 769 (2004), and R. Polk Wagner & Lee Petherbridge, Is the Federal Circuit Succeeding? An Empirical Assessment of Judicial Performance, 152 U. PA. L. REV. 1105 (2004). At any rate, the low influence of the Federal Circuit is a good example of how influence and reputation can diverge.

130. Gulati & Sanchez, supra note 4, at 1175; William M. Richman & William L. Reynolds, Elitism, Expediency, and the New Certiorari: Requiem for the Learned Hand Tradition, 81 CORNELL L. REV. 273, 319 n.220 (1996); see also White v. Samsung Elecs. Am., Inc., 989 F.2d 1512, 1521 (9th Cir. 1993) (Kozinski, J., dissenting from denial of rehearing en banc) (“For better or worse, we [the Ninth Circuit] are the Court of Appeals for the Hollywood Circuit.”); Marilyn F. Johnson et al., In re Silicon Graphics Inc.: Shareholder Wealth Effects Resulting from the Interpretation of the Private Securities Litigation Reform Act’s Pleading Standard, 73 S. CAL. L. REV. 773, 776 (2000) (suggesting that Ninth Circuit decisions on securities litigation will be significant given the presence in that circuit of Silicon Valley, home to “companies commonly targeted by attorneys bringing securities fraud class actions”).

131. Gulati & Sanchez, supra note 4, at 1176.

The simple longevity of the apparent high reputation of the Second and D.C. Circuits may also be an independent factor in their high reputation to date. In law, as in other walks of life, “[r]eputations die hard and are long in being born.” The modern Second and D.C. Circuits may, in part, free riding on their past reputational stock. (Incidentally, this makes the recent rise of the heretofore relatively undistinguished Seventh Circuit even more striking.) In other words, those circuits can be said to have developed a brand, or trademark, the durability of which influences subsequent snapshots of reputation. In the past two decades, the Seventh Circuit has developed its own brand.

Nonetheless, though I concede that the point is difficult to measure, my sense is that the reputation of all of the circuits has become homogenized in the latter part of the twentieth century. To put the same point somewhat differently, it is not so much that the reputations of the circuits have risen or fallen, but the entire notion of circuit reputation has become blurred and indistinct, at least in the eyes of some observers in the legal community. To the extent I am correct about the degradation of reputation in general, it is perhaps due in part to more lawyers, and more judges, trying to keep track of more judges and more cases.

When the circuits were first established in 1891, there were nineteen authorized judgeships. Congress has incrementally added positions since then. In the 1930s, during the heyday of Learned Hand’s service on the Second Circuit, there were a total of fifty-five judgeships, with the largest circuit having five and the smallest, three. As late as 1964, there were still only eighty-eight judgeships, and the largest circuit only had nine positions. Today, there are 179 authorized judgeships, with twenty-eight sitting in the largest circuit (the Ninth) and six in the smallest (the First). The average is twelve.

---


134. Cf. Landes, Lessig & Solimine, supra note 21, at 272 (discussing branding theory for individual judges).

135. For a summary of the information found in this paragraph, see the following table reproduced from the COMMISSION REPORT, supra note 46, at tbl.2-2:

<table>
<thead>
<tr>
<th>YEAR</th>
<th>LARGEST COURT</th>
<th>SMALLEST COURT</th>
<th>MODAL COURT</th>
<th>TOTAL JUDGESHIPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1892</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>1930</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>55</td>
</tr>
<tr>
<td>1950</td>
<td>9</td>
<td>3</td>
<td>6</td>
<td>75</td>
</tr>
<tr>
<td>1964</td>
<td>9</td>
<td>3</td>
<td>9</td>
<td>88</td>
</tr>
<tr>
<td>1978</td>
<td>26</td>
<td>4</td>
<td>11</td>
<td>144</td>
</tr>
<tr>
<td>1984</td>
<td>28</td>
<td>6</td>
<td>12</td>
<td>168</td>
</tr>
<tr>
<td>1990</td>
<td>28</td>
<td>6</td>
<td>12</td>
<td>179</td>
</tr>
</tbody>
</table>
(These numbers are somewhat misleading, since not all authorized positions are always filled, and such positions do not reflect circuit judges on senior status who still hear cases.\textsuperscript{136}) Recall the judge who said the Ninth Circuit was too big to possess an identifiable circuit identity.\textsuperscript{137} Perhaps the same observation holds true for many of the other circuits.

Consider, too, the actual or perceived evolution of characteristics of judges who make up the circuits. Some part of the reputation of the Hand generation of judges was their elite background and education. Yet the differences among judges of all of the circuits has seemed to moderate over time. To be sure, the Second and D.C. Circuits historically have had a high percentage of appointees who graduated from elite law schools—but the First and Ninth Circuits had even higher percentages.\textsuperscript{138} Likewise, despite recent attention focused on the subject,\textsuperscript{139} the federal judicial selection process has always been, at some level, “political” in nature. This is evidenced by the vast majority of appointees having affiliation with the party of the appointing President.\textsuperscript{140} There is also a perception that, especially on contentious issues like abortion, the death penalty, and civil rights, courts of appeals judges vote on a predictable, political basis—Republican appointees, conservative, Democratic appointees, lib-

\textsuperscript{136} For example, as of 1998, reportedly “when one includes senior judges and visiting judges, the true number of appellate judges is at least 266 (who are assisted on occasion by another 323 district judges).” Gulati & McCauliff, \textit{supra} note 117, at 172 n.64 (citing a statement by Professor Judith Resnik to the Commission on Structural Alternatives for the Federal Courts of Appeals); see also Resnik, \textit{supra} note 119, at 951 n.92.

\textsuperscript{137} See \textit{supra} note 77 and accompanying text. But see Richman & Reynolds, \textit{supra} note 130, at 301 (“[T]here is no empirical evidence that additional judgeships will reduce prestige or that reduced prestige will diminish the pool of judicial candidates.”).

\textsuperscript{138} See Susan Haire et al., \textit{An Intercircuit Profile of Judges on the U.S. Courts of Appeals}, 78 \textit{Judicature} 101, 102 tbl.1 (1994). The authors of the study took the list of elite law schools (that is, the top twenty-five) from the familiar rankings compiled by \textit{U.S. News & World Report}.


\textsuperscript{140} See SONGER ET AL., \textit{supra} note 132, at 29-45 (analyzing appointments to courts of appeals from Coolidge to Reagan); Sheldon Goldman et al., \textit{W. Bush Remaking the Judiciary: Like Father Like Son?}, 86 \textit{Judicature} 282, 308 tbl.4 (2003) (analyzing the same for Carter through the George W. Bush administrations).
eral. The perception is supported by some evidence, though it can easily be overstated. Most cases that come before three-judge panels are not highly contentious issues, and the large majority of such decisions are unanimous, no matter the ideological makeup of the panel. While the political nature of a position might attract some,

141. For an example from the press, see Deborah Sontag, The Power of the Fourth, N.Y. TIMES, Mar. 9, 2003, § 6 (Magazine) at 38, 40 (“[T]he Fourth Circuit is considered the shrewdest, most aggressively conservative federal appeals court in the nation.”).


As Frank Cross has observed, federal judges routinely deny that anything other than traditional legal principles guide their decisions. Frank B. Cross, Decisionmaking in the U.S. Circuit Courts of Appeals, 91 CAL. L. REV. 1457, 1464-67 (2003). He further suggests, “Judges may care about their reputation with the lawyers and litigants who appear before them. Dedication to the values of the judicial craft through conscious adherence to the legal model may enhance the judge’s prestige in the legal community.” Id. at 1474-75 (footnote omitted). My point is that to the extent it is perceived, rightly or wrongly, that many circuit judges vote based on their ideological values, rather than legal precedent, then in the eyes of many, their reputation, individually and collectively, will suffer. In this regard, it is perhaps significant that there is some statistical evidence that there was relatively less difference in voting between Democratic and Republican appointees on the circuits in the early part of the century. SONGER ET AL., supra note 132, at 114-16 (analyzing data from 1925 to 1988).

143. In their recent discussion of, and contribution to, the literature on decisionmaking by lower federal court judges, Greg Sisk and Michael Heise remark:

With respect to public policy, that judicial ideology may play a role at the margins in deciding certain types of controversial court cases cannot be gainsaid. But to suggest that partisan or ideological preferences are prevalent influences in deciding most cases or are invariably powerful variables in deciding even the most controversial and open-ended of legal issues is a dubious extrapolation from the empirical evidence.

Gregory C. Sisk & Michael Heise, Judges and Ideology: Public and Academic Debates About Statistical Measures, 99 NW. U. L. REV. 743, 746 (2005); see also Cross, supra note 142, at 1514-15 (indicating that in the empirical study of circuit decisionmaking, both legal model and judicial ideology play important roles in affecting judicial behavior); Sunstein et al., supra note 142, at 536 (“More often than not, Republican and Democratic appointees agree with one another, even in the most controversial cases.”).

A related point is that there is evidence that recent circuit court appointees are, at appointment, on average younger than previous appointees. Compare Goldman, supra note 54, at 354-56 tbl.9.2 (finding that the average ages of circuit appointees of FDR, Truman, Eisenhower, Kennedy/Johnson, Nixon/Ford, Carter, and Reagan were 52.9, 55.1, 55.9, 52.7, 53.4, 51.8, and 50.0, respectively), with Goldman et al., supra note 140, at 308 tbl.4 (finding that the average ages of circuit appointees of George H.W. Bush, Clinton, and George W. Bush were 48.7, 51.2, and 50.6, respectively). This presumably means that these appointees may have, on average, longer judicial careers. But it also means that their prestige suffers, or is more difficult to attain, since at least traditionally older judges may be considered more distinguished than younger ones. RICHARD A. POSNER, AGING AND OLD AGE 181-82 (1985). Exceptions immediately come to mind (consider Richard Posner, ap-
it may repel others. Finally, it is worth stating that the earning power of circuit judges is not lavish. They currently all earn $165,000, which is below many counterparts in the private sector, and increases in the twentieth century have only kept up with inflation.144 None of these points mean that circuit judges do not enjoy a lofty status in the legal profession. They simply suggest that, in the eyes of some, that status has dissipated over the past few decades.

The decisional output of the circuits can also be a factor in reputation being more diffuse. In 1930, less than 3000 cases were filed in the courts of appeals; as late as 1960 the figure was less than 4000, but it has risen rapidly since then. By the late 1990s, over 50,000 appeals were annually being filed.145 Thus, even though the number of judgeships has also risen, the average caseload for each judge has risen even faster. In 1960, for example, there were about 50 filings annually per judge, a figure that rose to 300 by the late 1990s.146 To be sure, not all filings result in an argued case and a decision on the merits, much less a published opinion thereafter. But the avalanche of opinions, both published and unpublished, produced each year surely makes it more difficult than in the past for anyone to follow and evaluate the collective work product of a circuit. Practitioners (and judges and their clerks) attempt to deal with the problem by using computerized databases for legal research. But that is not necessarily a panacea. Easier research makes it easier to cite more cases in briefs or opinions. This is not necessarily a bad thing, but it might suggest that the brand of a particular circuit opinion carries less weight than before.147

144. POSNER, supra note 36, at 21-33 (discussing extensively the history and present status of federal judicial salaries); Albert Yoon, Love’s Labor’s Lost? Judicial Tenure Among Federal Court Judges: 1945-2000, 91 CAL. L. REV. 1029, 1032-39 (2003) (same). Yoon points to evidence suggesting that recent judicial appointees are generally wealthier than before, which is perhaps a reaction to the failure of salaries to significantly rise. Yoon, supra, at 1056. On the other hand, federal judges enjoy significant perquisites, including control over their schedules, not needing to deal with clients or campaign contributors, the use of staff and facilities, and fairly generous retirement and pension payments. Id. at 1056-57.


146. Cooper & Berman, supra note 145, at 693 n.19.

147. Judge Bruce Selya of the First Circuit has made the following argument: Computer assisted legal research is much like laundry equipment in this respect. As cases remotely on point become even easier to find, the expectations for research rise, courts crank out more opinions, lawyers write more briefs (citing more opinions), and opinions cite more opinions. The cycle then begins anew. All too often, the judges are drained.
A related problem, in the view of some, is what has been colorfully labeled as the demise of the Learned Hand tradition. In brief, the argument runs that the overall quality of appellate opinions has diminished due in large part to the expanding-per-judge caseload, the delegation of opinion writing to law clerks or other circuit staff, and the increasing use of visiting judges to fill out many three-judge panels.\textsuperscript{148} It is not always clear what the critics mean by diminished quality. Some have pointed to the increase in recent years of longer, colorless, and heavily cited or footnoted opinions.\textsuperscript{149} The perception of a demise in quality of opinions surely has a negative impact on circuit reputation. There are significant exceptions to these generalizations. The D.C. Circuit, for example, typically publishes fewer opinions than other circuits\textsuperscript{150} and typically uses the fewest number of visiting judges.\textsuperscript{151} Nor can it be coincidence that many of the opinions of Judges Posner and Easterbrook, in particular, are considered in some quarters to be of higher quality than those of their peers.\textsuperscript{152}

Finally, changes in the size and shape of the audience may impact reputation. The legal profession as a whole has grown in size over the past century, has become increasingly specialized, and has become more diverse with respect to gender and minority status.\textsuperscript{153} Perhaps the typical practitioner has less need, desire, or leisure time to follow legal arguments. To make matters worse, the quality of legal argumentation sometimes seems to vary in inverse proportion to the rate of citation. Bruce M. Selya, \textit{Publish and Perish: The Fate of the Federal Appeals Judge in the Information Age}, 55 Ohio St. L.J. 405, 408 (1994). That said, an overload of cases may resurrect notions of reputation. Signals of prestige, such as the author of an opinion or the circuit from which it came, may be used as cues of quality, if there are so many opinions potentially to read and cite. Posner, supra note 9, at 68-69.


\textsuperscript{149} Posner, supra note 36, at 145-57; see also Gulati & Sanchez, supra note 4, at 1149 (“[T]he ‘best’ opinions are clear, concise, fully theorized, innovative, irreverent, placed in historical context, illustrative, and humorous. These are not the characteristics of the typical appellate court opinion.”).

\textsuperscript{150} Lindquist, supra note 49, at 28-29. The D.C. Circuit opinions, though, are not typically longer or more heavily footnoted than opinions from other circuits. Posner, supra note 36, at 155 n.53.


\textsuperscript{152} A market evaluation of this point is provided by the empirical study of Mitu Gulati and Veronica Sanchez, which demonstrated that the opinions of Posner and Easterbrook are disproportionately selected to appear in law school casebooks. Gulati & Sanchez, supra note 4, at 1155-56. The Seventh and Ninth Circuits led in opinions being utilized. Id. at 1155.

or be concerned with the reputation of general jurisdiction courts or of the judges that comprise them. This may be true—particularly regarding cases that are outside of the practitioner’s speciality. (On the other hand, the same phenomenon might accent the reputation of circuits in specialized areas—for example, administrative or environmental law in the D.C. Circuit and securities law in the Second Circuit.) In addition, it cannot be ignored that, until the recent past, white males constituted most judges on the circuits, regardless of the circuit’s reputation. Perhaps the concept of judicial reputation carries less weight for those members of the profession who are from groups traditionally underrepresented on the federal bench as a whole or among the great judges or circuits in particular.154

V. CONCLUSION

The reputations of judges and the courts they constitute ebb and flow, and those of the United States courts of appeals are no exception. There has been a good bit of discussion about the reputations of these courts but little effort to systematically examine them (as opposed to the burgeoning literature on the influence of one judge or court on another, often using citation analysis). I have attempted to fill some of this gap in the literature. But many questions remain, both backward-looking and forward-looking. Regarding the latter, further research could more closely examine the origins of the Second and D.C. Circuits’ high reputations, and whether such reputations were in fact commonly recognized in the pre-World War II era. Moving closer to the present, one might examine, say, the higher profile constitutional law and civil rights cases handled by the circuits since the 1950s to determine the effect on reputation. Regarding future developments, one can wonder how durable will be the reputation of the Seventh Circuit, when comes the day when Judges Posner and Easterbrook are no longer sitting. Will the much-maligned Ninth Circuit eventually have its reputation increase, perhaps (or perhaps not) at the expense of other circuits?155 What are likely reputational effects of the proposed breakup of the large Ninth Circuit into two or three new circuits?156 The difficulty of examining reputation will not prevent it from remaining an enduring subject for discussion.

154. I make the point with caution, since I am unaware of any rigorous study of this point. Likewise, I do not consider changes in other potential audiences, such as federal district judges or law school professors.

155. Cf. Jeff Chorney, 9th Circuit Dominates the High Court’s Docket, NAT’L L.J., July 5, 2004, at 6 (stating that the disproportionate number of Ninth Circuit cases reviewed by the Supreme Court may be explained by “the West [being] a cultural and economic powerhouse, a place where novel legal issues are simply more likely to come up”).

156. For a brief discussion of these proposals, see HART & WECHSLER, supra note 40, at 53 n.170. For further discussion, see Symposium, Managing the Federal Courts: Will the Ninth Circuit Be a Model for Change?, 34 U.C. DAVIS L. REV. 315 (2000); Special Issue on
A more basic line of inquiry should focus on the nature of reputation itself. Despite the historic preoccupation with status and prestige, perhaps those concepts are being deflated or changed, at least within some segments of our society, with regard to certain issues. Perhaps various shifts in politics, culture, business, and law have made reputation—as it was understood in the twentieth century—a less viable concept than in our century. 157 But these questions, too, must await further inquiry.

Studying and measuring judicial reputation can also be relevant to empirically measuring judicial performance, the subject of the symposium in which this Essay appears. Initially, the link may not be apparent. Measuring reputation statistically, as distinct from, say, measuring influence through citation analysis, may be considered so difficult that there is no way to include it as one metric of judicial performance, in a tournament of judges or otherwise. It is difficult to create precise statistical measures of judicial reputation. But that difficulty does not mean it is impossible to utilize reputation of a particular judge, or of a circuit as a whole, when considering judicial performance. Indeed, before the advent of citation analysis and other statistical measures, reputation, imprecise though it is, was the (or a) basis for gauging the performance of a judge or court. Perhaps surveys of judges and lawyers with regard to reputation could be included in measures of judicial performance.

Relatedly, scholars should consider the reputation of particular judges, and of multimember courts in general, as possible explanatory variables in judicial decisionmaking. Perhaps judges or courts that enjoy a high (or poor) reputation engage in opinion writing and decisionmaking in ways that are distinct—irrespective of statistical measures of influence as such. 158 With regard to multimember courts, incentives to improve or reward performance could or should be group-based—especially if appellate judge cooperation is a norm to be encouraged. In short, tournaments of judges could be court-based as well as judge-specific. In these ways, and perhaps others, judicial reputation as a distinct concept can play a useful role in the measurement of judicial performance.


APPENDIX: THE THIRTEEN FEDERAL JUDICIAL CIRCUITS
A TOURNEMENT OF VIRTUE

LAWRENCE B. SOLUM*

I. INTRODUCTION: THE MEASURE OF MERIT ........................................ 1365
II. WHAT IS JUDICIAL EXCELLENCE? .................................................. 1366
   A. The Problem of Disagreement About Judicial Excellence .......... 1367
   B. The (Mostly) Uncontested Judicial Virtues (and Vices) .......... 1368
      1. Judicial Incorruptibility and Judicial Sobriety ................. 1369
      2. Judicial Courage .................................................. 1371
      3. Judicial Temperament and Impartiality .......................... 1372
      4. Diligence and Carefulness ...................................... 1374
      5. Judicial Intelligence and Learnedness .......................... 1375
      6. Craft and Skill .................................................. 1375
   C. The (Mostly) Contestable Judicial Virtues ............................ 1376
      1. Competing Conceptions of the Virtue of Justice ............... 1377
         (a) Justice as Fairness ........................................... 1377
         (b) Justice as Lawfulness ....................................... 1380
      2. Competing Conceptions of Equity and Practical Wisdom ........ 1383
III. DISCERNING EXCELLENCE ........................................................ 1385
    A. Screening for Judicial Vice ........................................ 1386
    B. Detecting the Phronimos ............................................ 1386
    C. Recognizing the Nomimos ............................................ 1387
IV. THE MISMEASUREMENT OF VIRTUE ............................................... 1388
    A. The Case for Quantification ........................................ 1388
    B. Measuring the Wrong Qualities .................................... 1389
       1. Citation and the Rule of Law ................................... 1389
       2. Productivity and Carefulness .................................. 1393
       3. Fame Versus Excellence ......................................... 1394
    C. Gaming the Tournament of Judges .................................. 1394
       1. Gaming the Productivity Measure ................................ 1395
       2. Gaming the Citation Frequency Measure ......................... 1396
       3. Gaming the Judicial Independence Measure ..................... 1396
       4. Gaming Clerk Selection ......................................... 1397
    D. The Costs of a Gamed Tournament .................................... 1397
       1. Damage to the Rule of Law ...................................... 1397
       2. The Exclusion of Soft Variables ................................ 1398
       3. Decreased Transparency .......................................... 1398
       4. A Crisis ................................................................ 1398
V. CONCLUSION: THE REDEMPTION OF SPECTACULAR FAILURE .................. 1399

I. INTRODUCTION: THE MEASURE OF MERIT

How ought we to select judges? One possibility is that each of us should campaign for the selection of judges who will transform our own values and interests into law. An alternative is to select judges for their excellence—that is, for the possession of the judicial virtues: intelligence, wisdom, incorruptibility, sobriety, and justice. In an in-

* John A. Cribbet Professor of Law, University of Illinois College of Law. I owe thanks to both Loyola Marymount University's Loyola Law School and the University of San Diego School of Law for research support of work incorporated in this Essay and to Mitu Gulati for comments on an early draft.
fluential and provocative series of articles, Stephen Choi and Mitu Gulati reject both these options and argue instead for a tournament of judges—the selection of judges on the basis of measurable, objective criteria, which they claim point toward merit and away from patronage and politics.\(^1\) Choi and Gulati have gotten something exactly right: judges should be selected on the basis of merit—we want judges who are excellent. But Choi and Gulati have gotten something crucial terribly wrong: they have mistaken measurability for merit. A tournament of judges would be won by judges who possess arbitrary luck and the vices of originality and mindless productivity; the contest would be lost by those who possess the virtues of justice and wisdom. The judicial selection process should not be transformed into a game.

In Part II, “What Is Judicial Excellence?,” I tackle the tough problem that Choi and Gulati avoid—the explication of a theory of virtue for judges. In Part III, “Discerning Excellence,” I discuss how we can tell whether candidates for judicial office are bad, by which I mean incompetent, or are truly excellent. Part IV, “The Mismeasurement of Virtue,” engages the idea of quantitative measures of judicial performance as a proxy for excellence. Finally, in Part V, “Conclusion: The Redemption of Spectacular Failure,” I argue that Choi and Gulati’s idea is a rare and valuable thing—an idea that is both completely wrong and wonderfully illuminating. One more thing: Choi and Gulati focus exclusively on the selection of Justices for the United States Supreme Court,\(^2\) whereas my discussion will range a bit more broadly to include the selection of judges for other courts and tribunals.

II. WHAT IS JUDICIAL EXCELLENCE?

If our enterprise is judicial selection, the immediate question is, Who are the best or most excellent judges? Behind that question lies a more fundamental issue: What is judicial excellence? Stated differently, What makes one judge better than another? Choi and Gulati largely beg this fundamental question—focusing instead on particular metrics of judicial performance. In this Part, I will say a bit about why they beg the question and then attempt to remedy this defect in their work by sketching a theory of judicial excellence.

---


2. Choi & Gulati, *Tournament*, supra note 1, at 300.
A. The Problem of Disagreement About Judicial Excellence

It may well be the case that there is wide agreement that we should select excellent judges but disagreement about what counts as judicial virtue. The problem of disagreement about judicial excellence is one of the key starting points for Choi and Gulati's defense of empirical measurement of judicial excellence. Their strategy is to focus on a few criteria about which we can agree and which lend themselves to quantification. As they put the point:

While different visions of merit may exist, some are more widely held than others. Few would quarrel with the claim that a judge who displays productivity, intelligence, and integrity is better than one who does not.3

And Choi and Gulati readily admit that by making this move, their proposal would not provide a ranking of judges on the basis of merit, skill, or excellence.4 Their ambition is more modest—to do better than the status quo by providing some objective measure of judicial excellence:

Our simple measures do not provide a perfect metric for judging skill, but that is not the standard at which we are aiming. The goal is to demonstrate the availability of a set of objective measures for which we can easily collect data and analyze and that would better identify, at the outset, a merit-worthy pool of Supreme Court candidates.5

So far, so good. We have two assumptions. First, there is disagreement about the criteria for judicial excellence, and so we ought to seek judges who possess those aspects of excellence about which there is agreement. Second, of those criteria on which there is widespread agreement, only some lend themselves to quantification, and so a “tournament of judges” should focus on the criteria for judicial excellence that are measurable.

From a normative perspective, what seems quite odd about Choi and Gulati’s development of these ideas is that their analysis seems driven by the availability of data. That is, Choi and Gulati begin with the question, What aspects of judicial performance can we easily measure? Only after the measurability question is answered do they then ask, What qualities of good judging are the readily available metrics likely to measure? Of course, as a way of getting started, this method has much to commend itself. If one wants to conduct a tournament of judges, one must work with the data that is available. But getting started is one thing, and serious analysis is another. For us to

---

3. Choi & Gulati, Empirical Ranking, supra note 1, at 27.
4. Id.
5. Id. at 29-30.
take Choi and Gulati seriously, their analysis needs to be supplemented by another step—*the specification of the actual criteria for judicial excellence*.

Why is specification of the criteria for judicial excellence necessary? In order to determine whether a tournament of judges will improve judicial selection or make it worse, we need to know how the easily measurable aspects of judicial excellence relate to those that are difficult to measure. That relationship is crucial, because there is no a priori reason for ruling out the possibility that focusing on the measurable might have the unintended consequence of favoring judges with serious defects. If a tournament of judges were more than just pie in the sky and actually began to influence the judicial selection process, there is the further concern that an emphasis on measurable criteria as the determinants of judicial selection might actually make judges and their decisions worse rather than better. From both the theoretical and the pragmatic standpoints, an answer to the fundamental normative question—What makes for excellence in judging?—is essential.

**B. The (Mostly) Uncontested Judicial Virtues (and Vices)**

Choi and Gulati make an important point when they note that there is disagreement about the qualities that make for good judging. In recent years, quite a bit of judicial selection has largely been driven by the preference of political actors for certain outcomes on key issues (abortion, affirmative action, and so forth), and hence ideology has played a major role in judicial selection. Nonetheless, it may be possible to identify a set of judicial excellences on which there is likely to be widespread agreement.

Whereas Choi and Gulati work backwards, from measurability to virtue, we shall work forwards, starting with the notion of judicial virtue. By “virtue” I mean a dispositional quality of mind or character that is constitutive of human excellence, and the “judicial virtues” include both the human virtues that are relevant to judging and any particular virtues that are associated with the social role of judges. We begin with an account of those judicial virtues upon which we can mostly agree—which I shall call the “uncontested judicial virtues,” or more accurately, “the mostly uncontested judicial virtues.” “Uncontested” in this context reflects the notion that these virtues are based

---

6. With respect to ideology, judicial selection is arguably a zero-sum game. That is, pro-choice political actors (especially interest groups that focus on the issue) have little to gain from the appointment of a pro-life judge who possesses other fine qualities. And vice versa, pro-life political actors have little to gain from the appointment of pro-choice judges, even if they have many other virtues. Of course, abortion is not the only issue, but many such issues cluster together, which makes a simple left-right model of political ideology useful both analytically and empirically.
on noncontroversial assumptions about what counts as good judging and on widely accepted beliefs about human nature and social reality; the qualifier “mostly” reflects the fact that even an account of judicial excellence based on widely shared assumptions will be contested by some.

How can we get at the zone of agreement about judicial excellence? While there is a good deal of argument about which judges are the best, there is actually an astonishing amount of consensus about two related questions, Who are the very worst judges? and What are the worst judicial vices? No one thinks that the best judges are corrupt, drunk, cowardly, foolish, or stupid. This consensus suggests a strategy for articulating a theory of the uncontested judicial virtues. Let us begin with the worst judicial vices and identify the characteristics that are necessary to correct those defects. These characteristics will constitute the uncontested judicial virtues.

1. Judicial Incorruptibility and Judicial Sobriety

One judicial vice on which there is likely to be near universal agreement is “corruption.” Judges who sell their votes undermine the substantive goals of the law, because corrupt decisions are at least as likely to be wrong as they are to be substantively correct. Moreover, corrupt decisions undermine the rule-of-law values of productivity and uniformity of legal decisions and likewise undermine public respect for the law and public acceptance of the law as legitimate.

Almost anyone with common sense is likely to accept the conclusion that judicial corruption is a vice. If we accept judicial corruption is a vice, then what is the corresponding virtue? This question could become complex—because there are a variety of character flaws that might lead to corruption. One such flaw that may be an underlying cause of corruption is greed (or pleonexia)—because a desire for more than one's share (or entitlement) could lead a judge to accept bribes. All humans are at risk of mistaking wealth (which can only be a means) for a final end (something worth pursuing for its own sake). Some judges may resent the fact that they receive compensation that is sometimes only a fraction of that received by their peers in private legal practice—some of whom may be less talented.

We do not need to identify all of the possible vices that could lead to corruption in order to see that incorruptibility is an uncontested judicial virtue. There is no real controversy over the proposition that judges should be disposed to resist the temptations that lead to corruption. We call this disposition the “judicial virtue of incorruptibility,” even if it turns out that this virtue encompasses a variety of particular virtues each of which corresponds to a particular human vice that could lead to corruption.
There is another vice that is closely related to corruption but distinct from greed. Judges can become corrupted because their desires are not in order—because they crave pleasure or the status conferred by the possession of fine things. Judges, like the rest of us, can be corrupted by a taste for designer shoes, fast cars, loose companions, or intoxicating substances. More subtly, a judge could be corrupted by a desire for the finer things of life, for example, a magnificent home, the ability to confer lavish gifts upon one’s children, or the opportunity for luxurious travel.

Let us use some old-fashioned terminology and call the vice of disorderly desire “intemperance.” One might argue that intemperance is a purely private vice—that a judge’s preference for a third cosmopolitan, the latest from Jimmy Choo or Manolo Blahnik, or the company of good looking youthful companions is her own business and hence irrelevant to the question whether she is an excellent judge. Of course, a proportionate and well-ordered desire for such things is simply not a vice—or at least not an uncontested vice. But a disposition to disproportionate desires for such pleasures can lead to more than corruption. Most obviously, a judge who is intoxicated (or high) on the bench is likely to be prone to error. The disordinate pursuit of less-intoxicating pleasures can also impair judicial performance—by focusing a judge’s attention and energy away from judicial tasks.

There is a counterargument. It is a common human experience to have a friend, colleague, or acquaintance who is intemperate but nonetheless “gets the job done”—and who even performs brilliantly at times. Who has not encountered the lawyer who is a star by day but a lush in the wee hours or the friend whose life at work still holds together despite a drug problem? So, the argument goes, intemperance is not a judicial vice—at least not until it interferes with the performance of judicial duty. Even if the intemperate judicial candidate is a disaster at home, her intemperance should not disqualify her from judicial office if she performs at the office. This counterargument is ultimately unpersuasive. Of course, an intemperate judge can get lucky and “get away with it,” either appearing to do well or even actually doing well despite disordered desires. But in such cases “getting away with it” is a matter of luck; an intemperate judge is simply not reliable. A really damaging misstep is always just one cosmopolitan away.

The virtue that corresponds to the vice of intemperance could be called temperance, in the classical sense that encompasses the ordering of all the natural desires. But I propose that we use another term to refer to the judicial form of temperance. We have a saying that captures the intuitive sense that judges must have their desires in order: we say of a temperate human that she or he is “sober as a judge,” and this suggests that we name this virtue “judicial sobriety.”
2. Judicial Courage

Fear is one of the most powerful and familiar of the emotions. The disposition to feel too much fear makes us cowardly; the disposition to insufficient fear makes us rash. Courage represents a mean between cowardice and rashness. Let’s call the judicial form of this virtue “judicial courage.”

We might usefully subdivide the virtue of courage into two parts—which I shall call “physical courage” and “civic courage.” That judges need physical courage in order to be excellent as judges is a lamentable fact in many societies. We have recently been reminded of this fact by the tragic experiences of federal district judge Joan Lefkow, who was threatened by one defendant and whose husband and mother were murdered by a party to another case. A judge who could be intimidated by threats of physical violence could not reliably do justice in our relatively peaceful society—much less under conditions in which violence (or threats of violence) is even more prevalent, as may be the case where narcoterrorism or ethnic conflict is pervasive.

Judicial courage has a second dimension. Judges, like most humans, care about their reputations and social standing. Like the rest of us, judges seek the approval and companionship of their fellows. So, in addition to physical danger, judges may fear consequences of their actions that threaten status and social approval. This fear is dangerous because the law may require judges to make unpopular decisions. A judge who ordered school integration in the South might be shunned socially. In societies where the judicial branch wields significant power in cases involving hot-button issues (abortion, end-of-life disputes, and so forth), there will be occasions where doing what the law requires may be profoundly unpopular. For this reason, judges need the virtue of civic courage—the disposition to put the regard of one’s fellows in its proper place and to take it into account in the right way, on the right occasions, and for the right reasons. A judge with this virtue will not be tempted to sacrifice justice on the altar of public opinion. A civically courageous judge understands that the good opinion of others is worth having if it flows from having done justice and that social approval for injustice is an impermissible motive for judicial action.

3. Judicial Temperament and Impartiality

Like fear, anger is an emotion both familiar and powerful. Judges, like the rest of us, may be hot-tempered or cool and collected. And, like the rest of us, judges are likely to find themselves in situations where a hot temper could produce intemperate actions. This is especially true of trial judges, who are given the task of maintaining order in what may become emotionally charged circumstances. Litigants may ignore judicial authority or act with disrespect. Some lawyers may deliberately attempt to provoke the judge in order to elicit legal mistakes or “on the record” behavior that displays animus toward a party, which may serve as the basis for an appeal. In the face of such provocations, a judge with an anger management problem may “fly off the handle.” Intemperate judicial behavior may lead the judge to misapply the law or to distort the applicable legal standards in “the heat of anger.” Moreover, a hotheaded judge may become partial—pulling against the party who is the object of anger and displaying favoritism to that party’s opponent.

Aristotle identified proates, or “good temper,” as the corrective virtue for the vice of bad temper. In the judicial context, this virtue is so important that we have a phrase that expresses the virtue as a distinctively judicial form of excellence—“judicial temperament.” This phrase reflects our sense that the virtue of “good temper” is essential for good judging.

Is judicial temperament also required for judges who do not supervise trials? Appellate judges work in a cooler environment—provocative behavior by appellate lawyers is rare, although not unknown. The parties to an appellate proceeding frequently do not appear, and if they do, they sit in the audience without any formal participation in the appellate process itself. Some appellate courts proceed almost entirely on the basis of the briefs, dispensing with oral argument and hence with the opportunity for “live and in person” provocations. Nonetheless, good temper is essential for excellence in appellate judging. Appellate judges hear cases in panels or en banc—which create opportunities for friction among the judges themselves. Hot tempers can destroy collegiality and, with it, the opportunity for compromise and mutual understanding. Moreover, even a brief can elicit anger, and if anger becomes rage, it can have a blinding effect, depriving the judge of the ability to recognize the merits of an argument or a weakness in the judge’s own conception of the legal issues in a case.

If excessive anger is a vice, then what about its opposite? Is there a vice of deficiency with respect to anger? The Stoics are famous for answering this question in the negative; we might say that for the Stoics, the disposition to feel anger in any circumstances is a vice.9 The contrary view is that proportionate anger serves a valuable function—it alerts us to wrongs and motivates us to respond to them. A simple way of framing the issue is to ask which character from the 1960s television series Star Trek would make the best judge—Captain Kirk, Dr. McCoy, or Mr. Spock. Mr. Spock resembles the Stoic sage—he feels no anger and acts only on the basis of logic; we imagine Judge Spock reacting with equanimity to even the most severe courtroom provocations. Dr. McCoy is hot-tempered; we imagine him flying off the handle in response to outrageous behavior by the lawyer for a greedy corporation. Captain Kirk represents a mean between these two extremes; we imagine Judge Kirk as appropriately outraged by bad behavior and injustice but nonetheless remaining “in control,” and responding in an appropriate manner. The virtue of judicial temperament consists in having appropriate anger—anger for the right reasons, on the right occasions, and with a clear understanding of the consequences of its expression.

More concretely, when a party flouts the law or disrespects the participants in a legal proceeding, anger may be appropriate. Such appropriate anger alerts the judge to the existence of a situation that must be dealt with. In some circumstances, the judge will properly display such anger, giving a lawyer, party, or witness a stern warning. When someone persists in bad conduct, sanctions may be warranted; in such cases, giving an appropriate sanction is the right way to act on the basis of appropriate anger. But judges with the virtue of judicial temperament will not display their anger by ruling against an offending party on issues that are close or by exercising discretion on incidental matters so as to disfavor the anger-provoking party.

One reason that anger is an especially dangerous vice for judges is that anger can produce bias. For this reason, the virtue of judicial temperament is closely related to another judicial virtue, “judicial impartiality.” This virtue is a familiar feature of our conception of good judging. We want judges to be neutral arbitrators. A judge should be open to the law and evidence and not be biased in favor of one side or another. Such impartiality should extend not just to the parties but should also encompass the causes, movements, special interests, and ideologies that may be associated with those parties. When a judge takes the bench or lifts her pen to write an opinion, she

should put aside her allegiance to left or right, liberal or conservative, religiosity or secularism.

It is a mistake, however, to view impartiality as synonymous with disinterest. The virtue of impartiality is not cold-blooded. This is because the role of judge requires insight and understanding into the human condition. A good judge perceives the law and facts from a human perspective. Some facts are hot—charged with emotional salience. Some legal rules are morally charged—engaging our sense of indignation when juxtaposed with violative behavior. So the impartial judge is not cold-blooded; she is not indifferent to the parties that come before her. Rather, the judge with the virtue of judicial impartiality has evenhanded sympathy for all the parties to a dispute. When we say that impartiality is not indifference, we mean that the virtue of impartiality requires both sympathy and empathy without taking sides or favoring the legitimate interests of one side over those of the other.

4. Diligence and Carefulness

Judging is hard work, involving its share of drudgery. Some trials are long and boring. Some opinions require long hours of research and even longer hours of careful drafting. The temptation to shirk this work is accentuated by the fact that judges are not (and should not be) closely supervised. And the lack of supervision is compounded in jurisdictions that grant judges life tenure or long terms in office. It is hard enough to remove a judge for outright corruption; one doubts that any American judge has been removed on the basis of sloth alone. But slothful or lazy judges can do real harm. They are tempted to delegate too much responsibility to judicial clerks, substituting the judgment of the clerk for the judge’s own intellectual engagement with the case. Another temptation is to shape one’s decision in order to minimize one’s own workload. If granting the summary judgment motion takes a case off one’s docket, the slothful judge might grant the motion for that reason alone, sacrificing justice on the altar of expediency.

What is the virtue that corresponds to the vice of sloth? We might call it diligence. The diligent judge has the right attitude toward judicial work, finding judicial tasks engaging and rewarding. But more than a good attitude is required. An excellent judge must have an appropriate “energy level”—a product of both physical and mental health. The combination of these traits should translate into a judge who is capable of hard work when hard work is required. Such a judge will put in the required hours and sweat out the difficult tasks. Such a judge will not hesitate to make the right decision, even if that makes more work for the judge. Nowadays, encouraging settlements
may be an appropriate activity for judges, but a diligent judge will aim for just and efficient settlements and not for resolutions that serve the judge's own convenience.

Carefulness is closely related to diligence. No one can sensibly doubt that judicial carelessness is a vice. Careless decisions, careless drafting, careless research—any of these can lead to substantive injustice. Carefulness is especially important in the context of judging, because excellent judging frequently requires meticulous attention to details. The lazy judge may shirk the unpleasant task of mastering the structure of a complex statute or avoid the painstaking task of making sense of a tangled body of precedent. Likewise, it requires diligence and care to draft an opinion in which each and every sentence is worded with careful appreciation of the importance of precision and accuracy. An excellent judge has an eye for detail and a devotion to precision.

5. Judicial Intelligence and Learnedness

Can anyone doubt that stupidity is a judicial vice? All humans need intelligence to function well—but some tasks require more intelligence on more occasions. Judging is the kind of task that almost always requires smartness and sometimes requires extraordinary intelligence. Both law and facts can be complex. Only a judge with intelligence will be able to sort out the complexities of the rule against perpetuities or penetrate the mysteries of a complex statute. But more than intelligence is required. A truly excellent judge must also be learned in the law, because one cannot start from scratch in each and every case and because there is at least some truth to the notion that the law is a seamless web. In terms of the corresponding judicial vices, stupid and ignorant judges will be error-prone, likely to misunderstand and misstate the law, and unlikely to make findings of fact that are correct.

The need for judicial intelligence and learnedness is accentuated rather than diminished in an adversary system. It is true that good lawyering makes a judge's job easier; the lawyers can identify the relevant issues and call the judge's attention to the best arguments on each side of those issues that are in dispute. But in an adversary system, successful advocates will try to make a bad case appear better by deploying sophistry and rhetoric. Intelligent and learned judges can “see through” the obfuscation.

6. Craft and Skill

So far, our investigation has focused on what Aristotle called the moral and intellectual virtues. These are dispositions of character and mind that make for human excellence. Good judging requires
more than good character and intellectual ability. That is because judging includes elements of craft, and therefore a good judge must possess a skill set—the particular learned abilities that are to good judging what good bowing technique is to archery or good draftsmanship is to architecture. A full account of judicial craft is far beyond the scope of this Essay, but one particular aspect of judicial craft and skill demands attention. Excellence in judging (especially good appellate judging) requires particular skill in the use of language. Good judges must be good communicators. This aspect of judicial skill includes at least two parts—oral and written. It is obvious that trial judges need good oral communication skills; they must deliver a variety of oral instructions to the various participants in both trial and pretrial proceedings. Among these, jury instructions are particularly important. Written communication skills are especially important for appellate judges in a common law system, because of the doctrine of stare decisis. Because appellate opinions set precedent, a badly written opinion can misstate the law or state the law in a misleading way. A really well-drafted opinion, on the other hand, can clarify the obscure and illuminate the meaning of murky legal texts.

Good communication skills are also important to judges when they mediate between the parties to a dispute. A skilled judge can gain the trust and cooperation of the parties—resorting to the threat of sanctions only in those rare cases where force is truly necessary. In this way, good communication skills can increase the efficiency of judicial proceedings, allowing the judge to focus her attention on those issues and cases where settlement and cooperative processes are unavailing.

C. The (Mostly) Contestable Judicial Virtues

One advantage of a theory of judicial excellence is that it reveals a large zone of agreement. For all practical purposes, we can agree that judges should be incorruptible, courageous, good-tempered, diligent, skilled, and smart. But these (mostly uncontested) virtues do not tell the whole story about judicial excellence. Even if we agree in our judgments about who the very worst judges are—the corrupt, ill-tempered, cowardly, lazy, incompetent, and stupid ones—there are strong and persistent disagreements about who the best judges are. The partisans of Lord Coke may deride the accomplishments of Lord Mansfield; the admirers of Justice Brennan may be among the critics of Justice Scalia. This Part investigates the source of these disagreements about judicial excellence.

Once again, my strategy is to examine the judicial virtues. In particular, I shall argue that disagreements about judicial excellence are typically rooted in two disagreements about the nature of judicial
virtue. The first disagreement is about the nature of the virtue of justice. The second disagreement concerns the role of equity and practical wisdom. On the one hand, some disagreements about judicial excellence turn out to be disagreements about and within conceptions of the virtue of justice. On the other hand, further controversies about which judges are best hang on differences in the understanding of the role of practical wisdom in judging.

Although there are important disagreements about the virtues of justice and practical wisdom, there are certainly agreements as well. When stated at a high level of generality and abstraction, these virtues will command near universal assent. Almost everyone will agree that an excellent judge must be just (rather than unjust) and wise (rather than foolish). Let’s borrow the concept/conception distinction. We might say that there is agreement that the concept of the virtue of justice is required for judicial excellence, but that there is disagreement about which conception of the virtue of justice is best (or correct or most adequate). And likewise with the virtue of practical wisdom—we agree on the concept, but disagree about which conception of equity is the best one.

1. Competing Conceptions of the Virtue of Justice

What does the virtue of justice require? To answer this question, let us examine two different conceptions of the virtue of justice: justice as fairness and justice as lawfulness. (For short, I will use the phrases “the fairness conception” and “the lawfulness conception” to refer to these ideas.) I shall argue that conceptualizing the virtue of justice as fairness necessitates intractable disagreements about which judges are excellent, and that the competing conception, which emphasizes the idea that excellent judges are lawful, opens the door to agreement in judgments about who is just.

(a) Justice as Fairness

One influential conception of the virtue of justice is premised on the idea that the just and the lawful are separate and distinct. Of course, the view is not that all laws are unjust or that no just norms are law. Rather, the idea is that there is no necessary connection between legality and justice. If this were so, then the most plausible conception of the virtue of justice might be articulated as follows:

The Virtue of Justice as Fairness: A judge, \( J \), has the virtue of justice as fairness, \( V(j-f) \), if and only if \( J \) is disposed to act in accord with the best conception of fairness, \( F \), in situations, \( S \), where fairness provides salient reasons for action.

One might think that a judge who possessed \( V(j-f) \) would ignore the law altogether, but this is not the case. If this thought were correct, it would provide the basis for a devastating objection to the fairness conception—because it would require each judge to substitute her private judgments about what fairness requires for the duly enacted constitutions, statutes, and rules. Although I shall not provide the argument here, it seems plain that this would be a recipe for chaos.\(^{12}\)

But a defender of the fairness conception need not admit that a judge who acted on the basis of fairness would disregard the law entirely. Why not? Because the existence of legal norms will frequently give rise to considerations of fairness that will transform the moral landscape, creating salient reasons of fairness that motivate a judge who has \( V(j-f) \) to act in accord with the law. An example may help to clarify and illustrate this point. Suppose there is a dispute between Ben and Alice over Greenacre—a vacant and unimproved parcel of land. The law gives Ben title to Greenacre, which he has purchased, but Alice has begun to use Greenacre by planting a garden. In the absence of the institution of property law, it might be the case that Ben would have no claim on Greenacre—how would he acquire such a claim without some use or improvement of the land? But given the existence of property law, Ben would have a claim of fairness, because he has paid for Greenacre and has reasonably relied on the legal institution of property. If this is so, then the law has created a claim of fairness that otherwise would not exist, and a judge with \( V(j-f) \) would decide in favor of Ben—assuming, of course, that there were no other circumstances that created an overriding reason of fairness to decide in favor of Alice.

Nonetheless, the fairness conception faces a formidable objection because of the role that private judgment plays for judges with \( V(j-f) \). To articulate this objection, we need to highlight the distinction between two questions about fairness—which I shall call “first order” and “second order” questions of fairness. A first-order question of fairness is simply the question, Which action is fair given the circumstances? A second-order question of fairness concerns whose judg-

---

(discussing legal positivism and the separation thesis, but not in the context of the virtue of justice).

\(^{12}\) Of course, there may be some theorists who believe that judges do and should act on the basis of their sense of fairness rather than the law. Moreover, those who adhere to the radical or strong indeterminacy thesis contend that the law never constrains the choices of judges. See Lawrence B. Solum, On the Indeterminacy Crisis: Critiquing Critical Dogma, 54 U. Chi. L. Rev. 462 (1987).
ment about first-order questions will be taken as authoritative. Thus, the question—Given the fact of disagreement about the correct answer to a first-order question of fairness, whose judgment should be taken as authoritative?—is a second-order question of fairness. One possible answer to a second-order question of fairness is that one ought to rely on one’s own private judgment about what action is fair. A quite different answer is that one should rely on some source of public judgment. For example, one might rely on duly enacted and public laws.

The fairness conception implicitly requires judges to exercise private judgment about first-order questions of fairness. In exercising that judgment, the judge may conclude that expectations generated by reasonable reliance on the law provide reasons of fairness—as in the case of Ben, Alice, and Greenacre—but this is a conclusion of private judgment. One judge might conclude that Ben’s reliance on property law was reasonable and hence that fairness required a decision for Ben. A different judge might conclude that no one could reasonably rely on property law in cases in which they were allowing valuable land to lie fallow when others could make productive use of the land—and therefore decide for Alice. Yet a third judge might conclude that because of pervasive economic inequalities, the whole institution of property is unjust and award the land to a third party, Carla, who was in greater need than either Ben or Alice. Because each judge makes a private judgment about the all-things-considered fairness of following the law in each case, these judgments can (and we expect will) differ with the moral, religious, and ideological views of the particular judge.

The objection to the fairness conception of the virtue of justice is that disagreements in private judgments about fairness would undermine the very great values that we associate with the rule of law. Because the fairness conception requires each judge to exercise her own private judgment about what fairness requires—all things considered—and because such judgments will frequently differ, the outcome of disputes adjudicated by judges with \( V(j-f) \) will be systematically unpredictable. If this were the case, then the law would be unable to perform the function of coordinating behavior, creating stable expectations, and constraining arbitrary or self-interested actions by officials. How bad this would be is a matter of dispute. A Hobbesian answer to this question is that it would be very bad indeed—in the absence of a coordinating authority, life would be “solitary, poore, nasty, brutish, and short.”\(^{13}\) A Lockean answer is that reliance on

---

private judgment leads to “inconveniences,” but even an optimistic realist would surely concede that the inconvenience of a society that cannot secure the rule of law would be serious.

We are now in a position to apply what we have learned about the fairness conception to judicial selection. If the fairness conception were correct, then the excellent judges would be those who have the right beliefs about fairness and who are disposed to act on those beliefs. If we agreed on the content of the right beliefs about fairness, this would not be a problem, but we do not agree. So the fairness conception leads to disagreement about who has the virtue of justice. We can provide a crude translation of this point into the language of political ideologies of the left and right. For the left, only left-wing judges are just; because only left-wing judges have what the left considers true beliefs about what fairness requires. And, of course, for the right, the left-wing judges are unjust precisely because they have what the right considers false beliefs about fairness. Even the uncontested virtues—such as incorruptibility or courage—become problematic once the fairness conception has been accepted. For the left, an intelligent, diligent, and courageous right-wing judge may be worse than one who lacks a keen intellect, is somewhat lazy, and is susceptible to the pressures of public opinion. And for the right, these same concerns exist with respect to left-wing judges.

Another weakness of the fairness conception is that anyone who holds it is naturally tempted to apply a double standard of judicial excellence. The double standard works like this:

For judges with whom I agree, the fairness conception supplies the content of the virtue of justice. Right-thinking judges are excellent when they act on the basis of their convictions about what is fair. But when it comes to judges with whom I disagree, a different standard applies. Wrong-thinking judges are excellent when they stick to the rules. For them, the lawfulness conception provides the standard for the virtue of justice.

You may say that position is ludicrous; no one could hold such a blatantly inconsistent set of positions about the meaning of justice. In reply, I suggest that you pay careful attention to the political rhetoric that attends debates about judicial roles and judicial selection.

(b) Justice as Lawfulness

If the fairness conception of the virtue of justice is unsatisfactory, is there an alternative? In the Nicomachean Ethics, Aristotle suggests an alternative understanding of justice as lawfulness, but to
understand Aristotle’s view, we need to take a look at the Greek word *nomos*, which is usually translated as “law.” For the ancient Greeks, *nomos* had a broader meaning than does “law” in contemporary English. Richard Kraut, the distinguished Aristotle scholar, explained the difference as follows:

> [W]hen [Aristotle] says that a just person, speaking in the broadest sense, is *nomimos*, he is attributing to such a person a certain relationship to the laws, norms, and customs generally accepted by some existing community. Justice has to do not merely with the written enactments of a community’s lawmakers, but with the wider set of norms that govern the members of that community. Similarly, the unjust person’s character is expressed not only in his violations of the written code of laws, but more broadly, in his transgression of the rules accepted by the society in which he lives.

There is another important way in which Aristotle’s use of the term *nomos* differs from our word ‘law’: he makes a distinction between *nomoi* and what the Greeks of his time called *psēphismata*—conventionally translated as ‘decrees’. A decree is a legal enactment addressed solely to present circumstances, and sets no precedent that applies to similar cases in the future. By contrast, a *nomos* is meant to have general scope: it applies not only to cases at hand but to a general category of cases that can be expected to occur in the future.¹⁵

We can restate this last point by using our distinction between types of judgments (first- and second-order, private and public). If judges rely on their own first-order private judgments of fairness as the basis for the resolution of disputes, then it follows inexorably that their judgments will be decrees (*psēphismata*) and not decisions on the basis of a second-order public judgment—in other words, not on the basis of a *nomos*. A judge who decides on the basis of her own private judgments about which outcome is fair—all things considered—is making decisions that are tyrannical in Aristotle’s sense.

How can this be?, you may ask. Are not decisions that are motivated by fairness the very opposite of tyranny? But framing the question in this way obscures rather than illuminates the point. Of course, if there were universal agreement (or even a strong consensus) of first-order private judgments about fairness, then decisions on the basis of such judgments would be *nomoi* and not *psēphismata*. But our first-order private judgments about the all-things-considered requirements of fairness do not agree. So in any given case, a decision that the judge believes is required by fairness will be seen by others quite differently. At best, the decision will be viewed as a good-faith error of private judgment about fairness. More likely, those who disagree will describe the decision as a product of ideology.

personal preference, or bias. At worst, the decision will be perceived as the product of arbitrary will or self-interest. In no event will a decision based on a controversial first-order private judgment of fairness be viewed as the outcome of a *nomos*—a publicly available legal norm.

We are now in a better position to appreciate why rule by decree (*pséphismata*) is typical of tyranny. Decision on the basis of first-order private judgments about fairness is the rule of individuals and not of law. From Aristotle’s point of view, a regime that rules by decree does not provide the stability and certainty that is required for human communities to flourish.16 Kraut elaborates on this point:

> We can now begin to see why Aristotle thinks that justice in its broadest sense can be defined as lawfulness, and why he has such high regard for a lawful person. His definition embodies the assumption that every community requires the high degree of order that comes from having a stable body of customs and norms, and a coherent legal code that is not altered frivolously and unpredictably. Justice in its broadest sense is the intellectual and emotional skill one needs in order to do one’s part in bringing it about that one’s community possesses this stable system of rules and laws.17

And with that point in place, we can now formulate the lawfulness conception of the virtue of justice:

*The Virtue of Justice as Lawfulness*: A judge, J, has the virtue of justice as lawfulness, \( V(j-l) \), if and only if J is disposed to act in accord with the nomoi (positive laws and stable customs and norms), \( N \), in situations, S, where the nomoi provide salient reasons for action.

On the lawfulness conception, the virtue of justice does not require action in conformity with one’s first-order private judgments of fairness. Justice as lawfulness is based on a second-order judgment that judges (or, more generally, citizens) should rely on public judgments. The content of the public judgments are the *nomoi*—the positive laws and shared norms of a given community. Someone with the virtue of justice is disposed to act on the basis of the *nomoi*. In other words, the lawfulness conception holds that the excellent judge is a *nomimos*, someone who grasps the importance of lawfulness and is disposed to act on the basis of the laws and norms of her community.

We are now in a position to compare the fairness conception and the lawfulness conception. Which of these offers a more satisfactory conception of the virtue of justice? On the surface, it might appear that the fairness conception is more satisfactory—after all, who can

---

16. *Id.* at 106.
17. *Id.*
deny that we ought to do what fairness requires, all things considered? Although there is much more to be said in a full account of these matters, the argument advanced here provides good reasons to doubt that the fairness conception can offer a satisfactory account of the virtue of justice. A view of justice must take into account the distinctions between first- and second-order judgments and between public and private judgments. Once these distinctions are introduced, the need for second-order agreement on a public standard of judgment becomes clear. The lawfulness conception of the virtue of justice answers this need; the fairness conception does not.

2. Competing Conceptions of Equity and Practical Wisdom

But the virtue of justice may not be exhausted by the lawfulness conception. Even if we concede that in ordinary cases justice requires adherence to the law, the question remains whether there are extraordinary cases—cases in which excellent judges would depart from the law (or, to put it differently, decide that the law does not really apply). Even if first-order private judgment cannot do the work of filling in the content of a general conception of the virtue of justice, that does not necessarily imply that the judge’s sense of fairness has no role to play.

One reason we might doubt the adequacy of the lawfulness conception as the whole story about the virtue of justice flows from the fact that the positive law is cast in the form of abstract and general rules; such rules may lead to results that are unfair in those particular cases that do not fit the pattern contemplated by the formulation of the rule. If lawfulness were the whole story about the virtue of justice, then an excellent judge would apply the rule “come hell and high water,” even if the rule led to consequences that were absurd or manifestly unjust. But this implication of the lawfulness conception seems odd and unsatisfactory. Another way of conceptualizing this concern is to distinguish between two styles of rule application, which I shall call “mechanical” and “sensitive.”

Does the excellent judge apply the rules in a rigid and mechanical way? Or does a virtuous judge correct the rigidity of the lawfulness conception with equity? The classic discussion of these questions is provided by Aristotle in Book V, Chapter 10, of the *Nicomachean Ethics*:

What causes the difficulty is the fact that equity is just, but not what is legally just: it is a rectification of legal justice. The explanation of this is that all law is universal, and there are some things about which it is not possible to pronounce rightly in general terms; therefore in cases where it is necessary to make a general pronouncement, but impossible to do so rightly, the law takes account of the majority of cases, though not unaware that in this
way errors are made. And the law is none the less right; because the error lies not in the law nor in the legislator but in the nature of the case; for the raw material of human behaviour is essentially of this kind.\textsuperscript{18}

This is the \textit{locus classicus} for Aristotle's view of \textit{epieikeia}, which is usually translated as “equity” but can also be translated as “fair-mindedness.” As Roger Shiner puts it,

\begin{quote}
Equity is the virtue shown by one particular kind of agent—a judge—when making practical judgments in the face of the limitations of one particular kind of practical rule—those hardened customs and written laws that constitute for some society that institutionalized system of norms that is its legal system.\textsuperscript{19}
\end{quote}

But there is a problem with supplementing the lawfulness conception of the virtue of justice with the notion of equity. Understanding the problem begins with the fact that the virtue of equity seems to require the exercise of first-order private judgments of fairness. Once such judgments are admitted to have trumping force—to have the power to override the second-order judgment to rely on the public judgments embodied in the law—the question becomes how the role of private judgment can be constrained. Without constraint, private judgment threatens to swallow public judgment, and then we are on a slippery slope that threatens to transform the lawfulness conception into the fairness conception.

The trick is to constrain equity while preserving its corrective role. To put the point metaphorically, we need an account of equity that enables us to navigate the slope while providing sufficient traction to avoid slipping or sliding. An Aristotelian account of the virtue of equity gives us three points of traction. The first point of traction is provided by the distinction between the equitable correction of law's generality and the substitution of first-order private judgments for the \textit{nomoi}. Equity is not doing what the judge believes is fair when that belief conflicts with the law; rather, equity is doing what the spirit of the law requires when the expression of the rule fails to capture its point or purpose in a particular factual context. The second point of traction is provided by the virtue of justice itself. A judge who is \textit{nomimos} simply is not tempted to use equity to avoid the constraining force of the law. A \textit{nomimos} has internalized the normative force of the law; such a judge wants to do as the law requires.

\begin{flushright}
\end{flushright}
The third point of traction is provided by Aristotle’s understanding of the intellectual virtue of practical wisdom, or *phronesis*—think of the quality that we describe as “good judgment” or “common sense.”20 A judge with the virtue of practical wisdom, a *phronimos*, has the ability to perceive the salient features of particular situations. In the context of judging, we can use Llewellyn’s phrase, “situation sense,”21 or by way of analogy to the phrase “moral vision,”22 we can say that a sense of justice requires “legal vision,” the ability to size up a case and discern which aspects are legally important. The *phronimos* can do equity because she grasps the point of legal rules and discerns the legally and morally salient features of particular fact situations.

This account of equity can be contrasted with two rival accounts. On the one hand, we can imagine a conception of judging as pure equity—the idea that the judge would simply do the right thing in each particular fact situation. This conception of equity is simply a version of the fairness conception of the virtue of justice. On the other hand, we can imagine a conception of judging that limits equity to the vanishing point—perhaps to those cases where the application of the rule is truly absurd. Neither of these two alternatives offers a fully satisfactory account of the virtue of equity. The first alternative sacrifices the great goods created by the rule of law. The second alternative pays too high of a price for those goods and requires more rigidity than is necessary. A constrained practice of equity done by judges who are both *nomimos* and *phronimos* combines the values of the rule of law with the flexibility to bend the rules to fit the facts when that is required by the purposes of the rules themselves.

III. DISCERNING EXCELLENCE

Excellent judges possess the judicial virtues. They are incorruptible and sober, courageous, good-tempered, impartial, diligent, careful, smart, learned, skilled, just, and wise. But how can we tell which candidates for high judicial office possess these virtues? Knowing what judicial virtue *is* is one thing; knowing who possesses the judicial virtues is another. In this Part, I argue that the discernment of virtue has three components—screening for judicial vice, detection of wisdom, and recognition of lawfulness.

20. ARISTOTLE, supra note 18.
A. Screening for Judicial Vice

The first step in discerning excellence is the simplest. The initial screen for judicial excellence eliminates candidates who are incontrovertibly vicious—corrupt, ill-tempered, cowardly, unintelligent, or foolish. Screening for these vices is already a large part of the judicial selection process. Background investigations, conducted at the federal level by the FBI, seek to ferret out the moral vices. The solicitation of comments by peers (lawyers and judges) is designed to elicit evidence of more subtle defects in character or intellect.

If we want to effectively screen for vice, we want to select judges (and especially Supreme Court Justices) from candidates who have a track record that is likely to expose these vices. This suggests that Supreme Court Justices ought to be selected from judges or lawyers who have extensive experience in public life. Serious moral and intellectual defects may not be apparent at age thirty, but they are likely to have been exposed after two decades in public life. Luck may allow a cowardly or corrupt judge to flourish without incident for some span of years, but eventually vice will out.

B. Detecting the Phronimos

Mere absence of the worst vices is not enough. A good judge must possess the virtue of practical wisdom, or *phronesis*. Screening for vice is relatively easy; detecting the *phronimos* is likely to be both more difficult and more controversial. Before going any further, however, we ought to be careful not to exaggerate the problem. Practical wisdom is not an esoteric or mystic quality. Folk psychology recognizes “practical wisdom,” which is frequently called “common sense.” Our intellectual and literary traditions are full of references to the wise—from King Solomon to Gandalf. Our ordinary lives involve interactions with friends and colleagues whom we recognize as having good practical judgment; we ask them for advice and emulate their choices. Practical wisdom is harder to theorize than it is to recognize.

The fact that we are able to recognize practical wisdom offers the key to the problem of discerning the *phronimos*. Persons of practical wisdom, *phronimoi*, are recognizable by those who know them and interact with them. This fact has consequences for judicial selection. The process of selecting judges should rely heavily on the recommendations of those who are in a position to know whether the candidate possesses practical wisdom.

But this creates a special problem for the selection of Supreme Court Justices. The ultimate selector is the President, but the pool of

candidates is comprised mostly of judges. Given the separation of powers and the code of judicial ethics, judges may become cloistered—isolated from everyone but their friends and family, judicial colleagues, and law clerks. Opinions give evidence of craft and the intellectual virtues but provide an imperfect window on the judge’s practical wisdom. For this reason, it is especially important that judges—at least those who would be willing to serve on the Supreme Court—engage in practical activities that expose them to public life. Civic or charitable activities and service on judicial commissions are two obvious opportunities for judicial immersion in a public life of practical activity. Supreme Court Justices should be selected from among those who have demonstrated their possession of practical wisdom, both from the bench and in wider public life.

C. Recognizing the Nomimos

If judicial opinions are an imperfect window on the virtue of practical wisdom, they are well suited for the task of recognizing which judges are lawful and which are results-oriented. Although disregard for the rule of law can be masked by clever opinion writing, a persistent pattern of lawlessness is truly difficult to conceal. By way of contrast, a judge who is *nomimos* will strive to stay within the letter and spirit of existing law. Judges who believe in the rule of law attempt to give statutory or constitutional language its full due, eschewing interpretations that create unnecessary or artificial vagueness or ambiguity. Judges who believe in the rule of law will strive to follow precedent rather than evade it.

Of course, it might be possible for an ambitious lower court judge to feign the virtue of justice as lawfulness. But given the relatively small chance that any one judge has of appointment to the United States Supreme Court, it seems rather unlikely that many judges would choose to act as formalists when they are instrumentalists at heart. Judges with the vice of results-orientation are likely to wear it on their sleeve rather than conceal it underneath their robes.

In sum, we have good reason to believe that we can screen for vice, discern the possession of practical wisdom, and recognize true dedication to the rule of law. The fact that we have the capacity to recognize judicial virtue, however, does not entail that we can quantify it. A tournament of virtue, on the other hand, promises something that might appear to be a very great good. If we can quantify indicia of judicial excellence reliably, then judicial selection might proceed on the basis of objective, publicly available criteria. The hard question is whether the variables that can be quantified are good proxies for true judicial virtues.
IV. THE MISMEASUREMENT OF VIRTUE

Can we quantify judicial virtue? I will argue that the most reasonable answer to this question is “no.” Before I do, however, we should examine the case for quantification, as stated by Choi and Gulati.

A. The Case for Quantification

Choi and Gulati make the case for measurement by introducing the distinction between absolute and relative measurement of judicial excellence:

Some will see the search for a set of objective measures as pointless because they think that there is no way to measure or quantify what it means to be a good, let alone great, judge. This is likely true as an absolute matter. Nonetheless, with a set of candidates with track records as lower court judges, it may still be possible to make meaningful relative evaluations. So, just as it is impossible to articulate what special factor makes Lance Armstrong the best cyclist in the world, it is impossible to reduce Justice Benjamin Cardozo’s greatness as a judge to numbers. But one can look at how many times Armstrong has won the Tour de France and compare his numbers to those of his peers. Similarly, one can look at Justice Cardozo’s opinions and see how often they were cited by other judges, how often they were discussed in law reviews, and how often they made their way into casebooks. Justice Cardozo’s numbers can then be compared to those of his peers. As with Armstrong, this type of relative analysis does not give us a measure of his greatness or tell us what made him great. But it gives us a sense, even if imperfect, of how he performed relative to his peers.24

In other words, they argue that we can develop an ordinal scale for judicial excellence, even if we cannot develop a cardinal scale.

Before we go any further, however, we ought to observe that the analogy that Choi and Gulati make between bicycle racing and judging is a rather tenuous one. In the case of bicycle racing, there is an objective and quantifiable measure of performance. The first to finish is the winner; participants in the race are ranked (both cardinally and ordinally) by time. In racing, the output of the contestants is ultimately the time it takes each racer to finish, which can easily be compared across racers. In judging, there are many outputs: rulings, opinions, jury instructions, and so forth. These outputs cannot easily be compared across cases and judges. There is no scale that permits objective comparisons to be made.

B. Measuring the Wrong Qualities

Choi and Gulati's error extends beyond the obviously fallacious nature of their analogy to racing. The measures they propose—for example, citation rates and productivity—not only fail to capture the essence of judicial excellence, they may, at least in some circumstances, measure judicial vice. Choi and Gulati assume that judges who write lots of opinions that are cited a lot are better judges than those who write fewer opinions and get fewer citations. But are these assumptions correct?

1. Citation and the Rule of Law

Choi and Gulati seem to assume that citation rate correlates with judicial excellence. Their argument for this conclusion is actually somewhat obscure. It begins with the idea that there is a “market” for judicial opinions: “We can look at the frequency with which a judge’s opinion is used by a variety of consumers (including, for example, citation counts). Because circuit court judges write lots of opinions, the market test allows us to rank them in terms of the quality of those opinions.”25 In the accompanying footnote, they explain:

The view of judicial opinions as a market product available for consumption by judges, attorneys, and casebook writers has historical roots. In the early days of the Supreme Court, judicial opinions were typically recorded and distributed by private reporters. Reporters such as Cranch and Wheaton, for example, would record Court decisions, earning a return through private sales of their reports. Indeed, across the Atlantic in England, it was common for multiple reporters to record the same judicial opinion, competing against each other based on the quality of the text they provided.26

This passage does not, however, establish that the use of judicial opinions mimics a market in the respects that would be relevant to the notion that citation counts measure judicial excellence. The fact that reporters of opinions compete with each other on the basis of the accuracy of their reports does not entail that the authors of opinions compete with each on the basis of the quality of their decisions. Because this claim would be absurd, we can assume that Choi and Gulati did not intend to suggest it, and that the footnote is merely reporting an interesting fact and is not intended to establish the conclusion that there is a market for the excellence of opinions.

Choi and Gulati continue the development of their claim about the market for judicial opinions:

25. Choi & Gulati, Tournament, supra note 1, at 306.
26. Id. at 306 n.21 (citation omitted).
Markets do not always work well, however, and when they do not, they tend to be biased and inefficient. The problems that plague markets include asymmetric information, unsophisticated customers, and an inadequate number of producers (leading to oligopoly pricing).27

So far, so good. Choi and Gulati’s next claim, however, is problematic:

Unlike many other markets, however, the market for judicial opinions is relatively free of such imperfections. For one, judicial opinions may be obtained at no cost by judges and, in many areas of the law, are abundant.28

“No cost” is ambiguous. Choi and Gulati are right if they simply mean that judges do not personally pay a monetary price for access to law libraries and electronic legal databases. But this does not mean that the production of citations to other judges’ opinions is cost-free. Citing is an expensive business, but the price is paid in terms of time. Finding opinions takes time. Reading them takes time. Citing them properly takes time. Choi and Gulati obviously know this; they are legal scholars and personally pay a price for the citations they produce.

The price that judges pay in terms of time is an opportunity cost. Whether judges research, read, and cite on their own or have their clerks do this work, the time devoted to this activity is not available for other activities. Moreover, the time resource is finite. A judge’s own time is finite; there are only so many hours in the day, so many days in the week, and so many weeks in the year. Clerk time is also finite; judges are limited in the number of clerks they can employ. Typically, this limit is quite rigid, and even if a judge wanted to hire additional clerks, the judge would not be permitted to do so.29

The fact that citations are costly has important implications for answering the question whether citation rates measure the quality of judicial opinions. If citations were free, then one might suppose that the only variable that would influence the decision of Judge A to cite an opinion by Judge B would be the quality of Judge B’s opinion. Of course, there are other variables, such as whether Judge B’s opinions are binding on Judge A, but let us set those complications aside.

If, however, Judge A’s decision whether to cite Judge B’s opinion is costly, then quality will not be the only variable. Another important variable will be the costs of searching for Judge B’s opinion. If Judge B’s opinion turns up early in Judge A’s search for authority, then it will be more likely to be cited. As we all know, there are many

27. Id. at 306.
28. Id.
29. Some judges do have the discretionary power to hire “externs,” or law students who perform some of the tasks that clerks do.
basic propositions of law for which many possible opinions could be cited. In each federal circuit, for example, there are opinions on basic procedural matters (standards of appellate review, standards for summary judgment, and so forth) where hundreds or thousands of prior opinions will state the proposition of law.

A rational judge will not read all of these opinions and cite the opinion that does the best job of stating the law. Rather, the rational judge will read enough authority to be reasonably sure of the correct statement of the rule. When it comes to citations, we would expect judges to “satisfice” and not “optimize.” The opinions that are cited are likely to be the opinions the judge encounters first, and as a practical matter, this means that they are likely to be the opinions that result from traditional research methods. For example, a Westlaw search will produce a list of opinions in reverse chronological order. Each opinion that deals with the issue will cite other authority, and the recently cited authorities are highly likely to be cited in the newly written opinion. Importantly, judges are highly likely to cite authority that is already widely cited. If a particular authority is cited in the cases returned by the research process, there is an increased probability that the judge will cite that authority. And the more judges that cite the authority, the greater the likelihood that it will garner further citations.

Choi and Gulati continue their exposition, explicitly connecting citation rate with quality:

Indeed, the particular nature of the products (that they are free) means not only that competition is likely to occur effectively, but that we should be able to see clear and outright winners of the tournament. All judges will cite the best opinions. And to the extent certain “superstar” judges tend to write the best opinions, other judges will repeatedly look to these judges for guidance in the future. After all, given that the opinions all cost the same amount of money (zero), why not only use the best ones (even if the next best is only slightly worse)? This phenomenon of superstar judges does highlight one possible market defect: to the extent that most judges do not receive a large return from writing good opinions, many will not have an incentive to do so. All things considered, though, we predict that the reporting of objective ratings will raise the likelihood that more judges will exert effort to become a superstar judge (given the high payoff from winning the tournament).^30

But Choi and Gulati’s claim—that all judges will cite the best opinions—is clearly false once we look at citation through the lens of

---

30. Choi & Gulati, Tournament, supra note 1, at 307 (emphasis added).
networking theory. In the language of network economics,\textsuperscript{31} we can call this a process of “preferential attachment.”\textsuperscript{32} Opinions that are well situated in the network of citations will be cited many times; opinions that are more obscurely situated in the network will be cited rarely or not at all. The result is the so-called “rich get richer” phenomenon.\textsuperscript{33} Opinions that are initially cited for a proposition will be cited over and over for that same proposition.

In other words, the citation rate of a given opinion (and hence of the author judge) will depend in large part on the position of the opinion in the ecology of the network of authority. The first opinion to state a given proposition will be likely to generate many citations. Subsequent opinions will be more likely to be cited if they are the most recent opinion stating the proposition; the time that a given opinion remains at the top of the stack (the first position in the recency queue) will depend on the average frequency with which the proposition is stated at the time the opinion is issued. If the opinion of Judge C stating proposition X is at the top of the stack at a time when Judges D, E, and F are all working on opinions that also will state proposition X, then it is likely that Judge C’s opinion will be cited three times. But if Judge D states the proposition after Judge C and after Judges E and F have already finished researching their opinions, then Judge D’s opinion may never be cited at all. Moreover, once Judge C’s opinion has been cited by Judges D, E, and F, then it becomes highly likely that these judges will repeat their citations to Judge C’s opinion in future opinions. The repetition of the citation in their opinions increases the likelihood that other judges will cite Judge C’s opinion for the proposition. Occupying a very favorable node (or position in the ecology of the citations network) can result in an extraordinary number of citations; occupation of an unfavorable node can result in no citations at all. The important thing is that these differences can occur even though the proposition stated is exactly the same. For this reason, citation rates do not necessarily track quality.

But this understates the problem with citation rates as a proxy for judicial excellence. Given the ecology of citation networks, it seems quite likely that frequency of citation will be a function of originality. The first case to state a proposition is, all else being equal, highly likely to become an important node in the citation network. Whereas


\textsuperscript{33} See id. (discussing the rich-get-richer phenomenon in terms of hyperlinks).
judges must choose between many opinions when choosing authority for an oft-repeated proposition of law, they only have one choice when selecting authority for a novel proposition of law instantiated in only a single prior opinion. But it is hardly clear that novelty makes for good law or that originality is a judicial virtue. This is not to say that originality is never appropriate, but a truly virtuous judge will only be original when the law itself requires originality.

Indeed, I have argued that the opposite is true under normal conditions. The excellent judge is a *nomimos*, who follows the law rather than makes it. Good judges are clever in using the resources within existing law to solve the legal problems that come before them. The very best judges are experts at avoiding originality. And the very worst judges may be the most original. Very bad judges may use the cases that come before them as vehicles for changing the law, transforming the rules laid down into the rules that they prefer. This kind of results-oriented or legislative judging may produce many original propositions of law and hence a high citation rate, but this is a measure of judicial vice and not judicial virtue.

This is not to say that a high citation rate is necessarily an indicator of judicial vice. There are hard cases, in which some important issue of law comes before a court for the first time. Some judges may have high citation rates because the luck of the draw has handed them a disproportionate share of cases with truly new legal questions. But even if this is so, it does not follow that these are the best judges. Luck is not virtue.

2. *Productivity and Carefulness*

What about productivity? Choi and Gulati suggest that a tournament of judges should include a productivity measure:

The selection of a Supreme Court justice, therefore, should involve a prediction about the effort that a circuit judge is going to exert if elevated. Objective factors could focus on the effort that she exerted while she was a circuit judge. We could look at how many opinions (versus short form dispositions) the judge published, how many concurring and dissenting opinions she wrote, how many opinions she wrote in which she took on primary responsibilities (as opposed to delegating to clerks), and the overall number of cases which she played a role in deciding during a given period of time.34

But are the judges who write the most or longest opinions the best judges? Choi and Gulati have argued that short opinions are actually an indicator of judicial excellence, because shortness is a proxy for

34. Choi & Gulati, *Tournament*, supra note 1, at 309-10 (footnote omitted).
judges writing their own opinions as opposed to delegating that task to clerks.\textsuperscript{35} If total number of pages is not a good proxy for diligence, then what about the number of opinions written? It is certainly possible that the number of opinions written per time period is a proxy for judicial excellence, but this is not necessarily the case. The number of opinions written is surely a function of the number of opinions assigned. Assigning judges may attempt to equalize workloads; this might result in a judge who is given a difficult writing assignment being assigned fewer opinions. Or assigning judges might seek to equalize the number of writing assignments. The question whether there is a relationship between number of opinions written and judicial excellence seems to depend on a variety of empirical questions and not to be well suited for armchair speculation.

3. Fame Versus Excellence

There is a more general problem with Choi and Gulati’s approach to measuring judicial excellence. The judges who are cited most and who write the most opinions may well be the judges who want to be famous, or at least “almost famous.” Fame and glory (or external recognition) are powerful motivators, but it is not clear that a desire for fame is a virtue for judges. Indeed, the claim that excellent judges seek fame and glory seems somewhat counterintuitive.

There is nothing wrong with a desire for external recognition; humans as social creatures may naturally desire recognition by their fellows. But an excessive desire for fame is likely to be inconsistent with judicial virtue. The virtue of justice—the central component of judicial excellence—requires that judges aim at giving litigants what they are due, that to which they are entitled by the rules laid down. To the extent that judges decide cases on the basis of a desire for the fame and glory that come with winning a tournament of judges, they risk departing from the actions required by the virtue of justice; to put it more bluntly, a tournament of judges may create incentives to do injustice in order to win. Justice may require a prosaic opinion that says nothing likely to garner oodles of citations. Winning the tournament of judges may encourage a more dramatic opinion that makes new law in order to garner attention.

C. Gaming the Tournament of Judges

If there were a tournament of judges that influenced the selection of Supreme Court Justices, we may confidently predict that some judges would play to win. That is, they would view the tournament

as a tournament and devise strategies to maximize their chance of success. How might such a judge game the tournament of judges? To simplify, let’s assume the tournament is scored by a formula which includes the following three measures:

- Citations by lower courts, academics, and the Supreme Court.
- Productivity, including the number of majority and dissenting opinions written and the number of cases in which the judge participated.
- Judicial independence, or the frequency with which the judge is in opposition to another judge selected by the same President (or a President from the same political party).

Let’s also assume that the judicial selection tournament will be viewed by both participants and third parties as a game, with payoffs determined by the selection of Supreme Court Justices. Judges who are selected would receive a large positive payoff, but other players (Presidents, Senators, judges, academics, and law clerks) would also receive payoffs—if judges whose ideology they shared became Supreme Court Justices. This assumption, that judges would play to win, is shared by Choi and Gulati: “Our proposal also recognizes that judges, like the rest of us, respond to incentives.” So how would the judges respond to the incentives? How might the game be played?

1. Gaming the Productivity Measure

Choi and Gulati propose that we measure the number of opinions and dissents as well as the number of cases in which judges participate. How could this measure be gamed? Tournament leaders will wish to maximize the number of opinions and dissents. If not assigned an opinion, a judge will have a strong incentive to dissent. If two politically aligned judges sit on the same panel and one of the two is a tournament leader while the other is not, there will be a strong incentive to hand the opinion to the leader. Circuits determine their own procedures for case assignments. A circuit with a tournament leader who is politically aligned with the Chief Judge and the majority of the judges on the circuit will have a strong incentive to provide more opinion-writing opportunities to the leader. This will advantage judges in friendly circuits and disadvantage judges in unfriendly circuits. In the long run, however, there are only so many

---

37. Choi & Gulati, Tournament, supra note 1, at 305-09.
38. Id. at 308-10.
39. Id. at 310.
40. Id. at 305.
ways to game productivity. Presumably, equilibrium will be reached among the judges who are tournament leaders—with each scoring in approximately the same range on this measure.

2. Gaming the Citation Frequency Measure

The opportunities for gaming this measure are obvious. Academics will now have an incentive to cite their favorites to influence tournament results. Likewise with both lower court judges and Supreme Court Justices. A set of second-order tactics will be likely to emerge. The composition of law school faculties can be influenced by state legislatures and by the wealthy alumni of private universities. The lower federal court benches are selected by the President and the Senate. Moreover, judges themselves can change their opinion writing so as to maximize the opportunities for both citing other judges (allies in the tournament) and for being cited. Opinions will become longer and long string cites will become the rule. Basic and uncontroversial issues will be discussed in depth. When faced with a choice between writing an opinion on an issue where there is no law—because the issue arises infrequently—and an issue on which there is a lot of law—because the issue comes up all the time—the rational tournament participant will avoid the former and seek the latter.

Perhaps the most successful tactic for gaming the citation frequency measure is also the most problematic. Judges will have an incentive to change the law, because an opinion that makes new law—especially new law on a topic that arises frequently—is much more likely to be cited than an opinion that merely restates existing law.

3. Gaming the Judicial Independence Measure

Choi and Gulati propose that we measure independence by voting records.41 Judges would score points for voting against a judge appointed by a President of the same party as appointed that judge. There are several ways to game this measure. The most obvious way is to dissent when a same-party judge is in the majority and the decision would otherwise be unanimous. And, by the way, the judge trying to win the tournament will also write a long, citable dissent that rehearses all of the basic law surrounding the case and cites all the judge’s allies in the tournament. Of course, there will be cases in which the players cannot decide contrary to party affiliation without changing the outcome. But if you are a tournament leader and the case is not on a hot-button issue about which you care deeply, it may well be in your interest to score some independence points by decid-

41. *Id.* at 310.
ing the case in a way you believe is wrong—and by writing a long opinion, of course!

4. Gaming Clerk Selection

Getting really good clerks is going to be very important in the tournament of judges. If you want to be a tournament leader, you will need to write a lot of very long opinions and dissents. Moreover, you need high-quality opinions, because they are more likely to be cited by other judges. So you want the best clerks. Supreme Court Justices can influence who gets the best clerks by informally signaling that some judges are “feeder judges.” Clerks will want those clerkships, because they will lead to prestigious Supreme Court clerkships, which in turn will lead to prestigious academic positions, creating the opportunity to influence both citations and future clerks. The advantage added by the very best clerks is likely to be substantial, and may well be decisive, given that citation frequency is the one measure among the three where an equilibrium ceiling is unlikely to be established by the players. With great clerks, a stable of externs, and some high-quality politicking, it might be possible for a judge to garner many thousands of citations.

D. The Costs of a Gamed Tournament

In their original article, Choi and Gulati suggested that the tournament of judges could be accompanied by a ban on discussion of any other merit-based criteria for judicial selection other than the tournament results. When it comes time to select Supreme Court Justices, the tournament results will be the only information that Presidents and Senators may use to justify their decision—other than political ideology. Assume that the tournament does, in fact, determine who is appointed to the Supreme Court. What price would we pay?

1. Damage to the Rule of Law

One thing that is very difficult to measure objectively is whether a judge has decided in accord with the law—rather than on the basis of either ideology or to gain an advantage in the tournament. The virtue of justice is not rewarded in the tournament. No points are assigned for getting the law right. Moreover, too high a regard for justice is likely to be punished. Judges who vote based on the merits will lose opportunities to write opinions and dissents. Judges who agonize about getting it right will be diverting precious time from the oppor-

42. See id. at 313 (“To address the problem of political transparency, an extreme form of the tournament would be one that bars the president and the Senate from putting forth merit-related rationales outside our list of objective factors.”).
portunity to score points by getting it long, that is, producing lots of long and citable opinions. And judges who get it right are unlikely to produce opinions with lots of novel propositions of law—and hence lots of citations.

2. The Exclusion of Soft Variables

Practical wisdom, or phronesis, is a key component of judicial excellence, but the tournament of judges does not award points to judges who have common sense, which is the ability to size up a situation and penetrate to the issues that are truly important. Indeed, the judges who possess this virtue are likely to be rather weak performers in the tournament of judges. They are likely to perceive that scoring points at the expense of doing justice is a rather poor excuse for judging. They are likely to lag behind their more canny and competitive colleagues.

3. Decreased Transparency

Choi and Gulati claim transparency as an advantage for the tournament of judges, but the opposite may be the result of their proposal. The tournament is likely to create an illusion of objectivity. Behind the scenes, however, there would be manipulation of opinion counts, citation counts, and independent decision counts. This will especially be true if one party controlled the Presidency, the Senate, the Supreme Court, and a majority of court of appeals slots at the beginning of the tournament. That party would have enormous strategic advantages in gaming the tournament, but the political nature of the selection process would effectively be masked by the apparently neutral and objective basis that the tournament results would provide for the selection of Supreme Court Justices.

4. A Crisis

The end result of Choi and Gulati’s proposal would be so awful that one cannot imagine the story ending except in some kind of crisis. You may not like the current Supreme Court, but imagine a court populated by judges who had won Choi and Gulati’s tournament. These judges would be without the virtues of integrity, wisdom, or justice. They would have been selected for the ability to manipulate the tournament results. In order to do this, the winning judges would be those who are willing to elevate self-interest over the interests of the public and the parties who appear before them. And these clever but vicious judges would be entrusted with the ultimate constitutional authority.

43. Id.
V. CONCLUSION: THE REDEMPTION OF SPECTACULAR FAILURE

If viewed as a serious proposal for reform of the judicial selection process, Choi and Gulati have a spectacularly bad idea—a real stinker. This can be true even if the retroactive application of Choi and Gulati’s selection criteria identifies excellent judges. The reason for this is obvious. No one had an incentive to game Choi and Gulati’s hypothetical tournament. The participants could not predict that the tournament would exist, and even if they knew that two law professors were conducting a hypothetical tournament, they would have very little incentive to play to win. All of this would change if the tournament of judges were actually implemented. Choi and Gulati recognize the imperfections of their measures of excellence and offer the following defense:

We will never succeed in generating a perfect objective measure of judicial quality. The point, however, is not whether objective criteria perform better than a perfect system of judicial selection. Rather, the question is whether objective criteria work better than the selection process we have today. Given how politicized the selection of Supreme Court justices currently is, the use of any objective factors will lead to a marked improvement.44

But this argument fails on two counts. First, it is not necessarily the case that the use of objective factors would lead to a “marked improvement” in a gamed tournament. Such a tournament would select for those who are motivated by a desire to win the tournament and not by those who want to do justice. Choi and Gulati assume that the tournament will reduce the role of political ideology, but in a gamed tournament, that assumption is doubtful. Second, the reform of the judicial selection process—like most reform processes—is likely to involve path dependency and opportunity costs. If political capital were invested in a tournament of judges and we start down the road of judicial selection based on objective measurement of outputs, then it may become more difficult to focus on true judicial excellence. If the tournament of judges favors one ideological faction over another, the winning faction will have every incentive to preserve the tournament. The opportunity cost of a real-world tournament of judges could well be loss of the chance for the implementation of real merit-based judicial selection.

Sometimes, however, bad ideas spark good debates. We can view Choi and Gulati’s tournament of judges as a thought experiment rather than a proposal for reform. As a thought experiment, the tournament of judges is a marvel, precisely because it invites rigorous analysis of the judicial selection process. In the end, Choi and

44. Id. at 312.
Gulati’s tournament of judges invites us to ask two questions: What constitutes judicial excellence? and How can we select judges who possess them? Those questions are worth answering.
INFORMATION AND THE SELECTION OF JUDGES: 
A COMMENT ON “A TOURNAMENT OF JUDGES”

AHMED E. TAHÀ*

I. INTRODUCTION ................................................................. 1401
II. EFFECT ON THE SELECTION OF JUDGES ................................. 1403
III. EFFECT ON JUDICIAL BEHAVIOR ........................................ 1407
IV. A TOURNAMENT OF U.S. DISTRICT COURT JUDGES? ...................... 1412
V. CONCLUSION .............................. .............................. 1413

I. INTRODUCTION

Choi and Gulati’s proposal for a tournament of judges,1 with appointment to the U.S. Supreme Court as the prize, has generated a furor and, more importantly, this law review Symposium. In response to an earlier version of their article, I raised a number of concerns with their proposal.2 In this Essay, I will expand upon some of these concerns and discuss how Choi and Gulati address them in the final version of their article.

Choi and Gulati correctly view the nomination and confirmation of Supreme Court Justices as driven largely by politics.3 Although they do not oppose politics playing some role in the selection of judges,4 they object to arguments about a judge’s political views being disguised as arguments about a judge’s nonpolitical qualifications, which they refer to as a judge’s “merit.”5 They argue that such mischaracterization delays the confirmation process by allowing political battles to be fought under a more respectable, nonpolitical pretext and unfairly tarnishes the reputation of nominees by encouraging insincere objections to their merit.6

To solve this problem, Choi and Gulati present an interesting proposal: all judges on U.S. courts of appeals will receive a ranking that reflects their merit.7 The ranking will be based solely on a number of quantitative, objective criteria that are correlated with a judge’s merit, such as the number of opinions the judge has published and

---

* Assistant Professor of Law, Wake Forest University School of Law. J.D., Ph.D., Stanford University, 1996; B.A., B.S., University of Pennsylvania, 1988. I am grateful to Michael Green, David Logan, Alan Palmiter, and Ronald Wright for helpful comments.
2. Some of these concerns were summarized in my contribution to a Jurist online symposium on the judicial confirmation process. See Ahmed E. Taha, Information and the Selection of Judges, Kurist, at http://jurist.law.pitt.edu/forum/symposium-je/choi-gulati-orth-taha.php (Apr. 15, 2004).
3. Choi & Gulati, supra note 1, at 299.
4. See id.
5. Id. at 301-02.
6. Id. at 299.
7. Id. at 303.
the number of times his or her opinions are cited by other judges or by legal scholars.\textsuperscript{8}

Of course, if those involved in the judicial nomination and confirmation process still wished to support or oppose a Supreme Court nominee for political reasons, they could nonetheless use the guise of a merit-based argument by claiming that the ranking is a flawed measure of merit. They will allege that a more accurate measure would have led to a different conclusion regarding the nominee’s merit and will claim to support or oppose the candidate on such alternate merit-based grounds, although they are truly motivated by political considerations. To prevent such subterfuge, Choi and Gulati propose—as an extreme version of their proposal—that the President and Senators not be permitted to put forth merit-based rationale other than the ranking.\textsuperscript{9}

They suggest, as more modest reforms, simply publicizing judges’ merit rankings and/or the judges’ scores in the criteria that underlie the ranking.\textsuperscript{10} They argue that this publicity would encourage the use of such factors in nominating and confirming judges and would pressure the President and Senators to provide an explanation when their support of, or opposition to, a particular nominee does not conform to such objective measures of merit.\textsuperscript{11}

In addition, Choi and Gulati argue that a ranking system would encourage desirable judicial behavior.\textsuperscript{12} Judges who have Supreme Court aspirations will seek to improve their rankings,\textsuperscript{13} and even judges who are not motivated by such hopes will seek high rankings to build their reputations as good judges.\textsuperscript{14}

Although I do not believe that the current judicial selection process should be replaced with a tournament, Choi and Gulati’s proposal is valuable because it encourages discussion focused on identifying the characteristics of a good judge and how to measure those characteristics. In addition, when refined, such measurements could provide useful information to persons involved in the judicial nomination and confirmation process. Nevertheless, when it comes to a tournament, the devil may be in the details. Although Choi and Gulati believe that even a flawed tournament would be better than the current selection process,\textsuperscript{15} a seriously flawed ranking system could

\begin{footnotes}
\item 8. Id.
\item 9. Id. at 313.
\item 10. Id. at 322.
\item 11. See id. at 312.
\item 12. Id. at 313-15.
\item 13. Id. at 314-15.
\item 14. Id. at 313-14.
\item 15. Id. at 304.
\end{footnotes}
create undesirable incentives for judges and could actually make the judicial selection process, and the judiciary itself, more political.

In this Essay, therefore, I wish to discuss some reservations regarding the tournament proposal. These fall under two broad categories: the effect of rankings on the judicial nomination and confirmation process and the effect of rankings on judges’ behavior. In addition, I argue that the case for a tournament of judges is stronger for positions in U.S. courts of appeals than on the U.S. Supreme Court. Federal appellate judges often come from the ranks of federal district judges, and a ranking of federal district judges would avoid many of the problems that might be created by ranking appellate judges. In addition, federal district judges are more likely than federal appellate judges to respond to the incentives created by a tournament.

II. EFFECT ON THE SELECTION OF JUDGES

The effect of a ranking system on the judicial nomination and confirmation process depends upon the extent to which the President and Senate take the ranking seriously. If the ranking is perceived as a largely flawed measure of judicial merit, it is unlikely to influence significantly the selection of judges, even if citing alternative measures of merit is prohibited. Opponents of a particular judge’s ranking will point out the ranking system’s serious defects and may persuasively argue that the ranking should be ignored.

Thus, Choi and Gulati’s proposed gag rule on alternative merit arguments makes it even more important that the ranking be well regarded by those persons involved in the nomination and confirmation process. If the ranking is perceived as a very weak measure of merit and the President and Senate are prevented from pointing to other indicators of merit, the selection process might place even less weight on judicial merit than it does currently.

Ultimately, then, the effect of a ranking system may depend upon how accurately the ranking is perceived to measure judicial merit. Of course, any ranking based on objective measures will be imperfect. It is beyond the scope of Choi and Gulati’s article to specify exactly which criteria should be included in the ranking and, as important, the weight that each criterion should receive in calculating the ranking. Nevertheless, they discuss several criteria that might be included in the ranking, including a judge’s publication rate, the number of citations by other judges and by legal scholars to those decisions, and how often the judge is reversed. However, an analysis

16. Id. at 303.
17. Id.
18. Id. at 307.
of these criteria demonstrates the limitations of any ranking methodology.

As Choi and Gulati note, the fact that a U.S. court of appeals generally decides cases in panels of three judges limits the usefulness of any objective measure of the quality of a particular judge, including the number of opinions the judge publishes. For example, deciding which panel member will write the majority opinion—and whether an opinion should be written or published at all—is a group decision. Thus, the number and topics of opinions that a judge publishes and the resulting total number of citations to a judge’s opinions are partly determined by the publication preferences of other judges in the circuit.

Intercircuit comparisons of judges are less meaningful than comparisons within a circuit. Each U.S. court of appeals has adopted its own criteria for the publication of opinions, and these criteria may affect the number of opinions judges in that circuit publish. Intercircuit differences in the number and types of cases filed might also affect the number of cases a particular judge hears or the judge’s publication rate. The total number of citations to a judge’s opinions should also be higher if the judge publishes more opinions. Also, the effects of deciding cases in three-judge panels likely differ across circuits. For example, a judge in a circuit where other judges disfavor frequent publication of opinions may publish less than she would if she were in a circuit where the other judges favor publishing.

The difficulty of intercircuit comparisons is not fatal for a ranking system. For example, a criterion in the ranking could be how many opinions a judge has published, relative to the average number of published opinions by other judges in the same circuit. However, that adjustment would discount the effort of judges in circuits where publication is more favored. Although such issues do not mean that a ranking will not contain valuable information regarding judicial merit, they do suggest that any objective measure of merit will be subject to valid criticisms. Such criticisms would certainly be made by those persons in the judicial selection process who support a judi-
cial nominee with a low ranking or who oppose a nominee with a high ranking.

Unfortunately, even if an objective ranking is constructed that fairly and accurately measures merit, it might not significantly affect the judicial selection process. Most of the ranking criteria that the authors suggest are already easily accessible. For example, the President’s or Senators’ staffs already can search Westlaw or LEXIS to calculate the number of opinions a judicial nominee has published and the number of citations to those opinions. Indeed, a judge’s opinions are already read as part of the selection process. The fact that neither the number of opinions a judge publishes nor the number of citations to those opinions is currently used in the debate over nominees indicates that those involved in judicial selection do not believe these measures are relevant to the selection decision. Thus, we should not be confident that they will pay attention to rankings that are based on such criteria.

More generally, all participants in, and most observers of, the judicial selection process know that politics play a large role. The President rarely nominates candidates outside of his party, and when there is significant disagreement in the Senate about a candidate, the vote splits largely on party lines. Also, the vocal interest groups that get involved in the selection process are typically seeking a judge who supports their positions on particular political issues. The fact that the President has the opportunity to fill any Supreme Court vacancies that arise has even been a campaign issue in recent presidential elections.

This raises another important limitation of a ranking system. A main purpose of Choi and Gulati’s proposal is to encourage honesty and transparency. No longer could preferences regarding a judge’s political or judicial ideology be hidden behind merit-based arguments. However, the ranking system will not reduce some of the major related problems that they identify.

Choi and Gulati lament the primary role politics plays in the confirmation process and point out the frequent hypocrisy exhibited by those involved in the process. They endorse Erwin Chemerinsky’s observation that when it is in a preferred candidate’s interest, many participants in the confirmation process switch their own stated posi-

24. For example, two recent controversial nominees to U.S. Courts of Appeals—Texas Supreme Court Justice Priscilla Owen and U.S. District Court Judge Charles Pickering—were rejected by the Senate Judiciary Committee on party-line votes. Helen Dewar, Senate Panel Rejects Bush Court Nominees, WASH. POST, Sept. 6, 2002, at A1.
25. Id.
tions regarding whether a nominee’s political views are relevant. 27 This inconsistency may contribute to the cynicism that many Americans have about their elected officials and about the judicial selection process.

However, a ranking system will not reduce this problem. Even in its most extreme form—which prohibits other merit-based arguments—their proposal will prevent only disputes about a judge’s merit. Participants in the confirmation process will still argue about—and take inconsistent positions regarding—the relevance of a nominee’s political views. In addition, there will still be intense debates about whether a judge actually has extreme political views.

Although there might be some increase in transparency by forcing such political debate from behind a “merit” mask, this effect may not be great. Indeed, as the debates surrounding Chief Justice Bird and Judge Bork illustrate, politicians are willing to make explicitly political arguments regarding a judicial nominee. This also suggests little cause for optimism that a ranking system would cause the judicial nomination and confirmation process to take place more rapidly. All the parties involved in confirmation battles likely know they are principally battles about politics, not about merit, even if they are sometimes phrased in merit terminology. Although a ranking system may make these confirmation battles more explicitly about politics, these battles will probably remain as intense and protracted as they are currently.

Ironically, if a ranking system does increase the emphasis that the confirmation process places on a judge’s merit, it might have another undesirable effect: it may allow a President to nominate more politically extreme candidates who happen to have high merit rankings. This could lead to an even more protracted and heated confirmation process, as well as a more politically extreme or splintered Supreme Court.

27. Id. at 322 n.6. Specifically, Chemerinsky observed that [in California, in 1986, conservatives argued that [Chief Justice Rose] Bird, and two other Justices . . . should be rejected because of their liberal views and prior votes, especially in death penalty cases. Liberals in California argued that assuring judicial independence required that evaluation be limited to the justices’ competence; that the individual’s ideology and prior votes should play no role in the retention process. But the sides were reversed a year later in a battle over the [Judge] Bork confirmation. It was the liberals who argued that Bork should be rejected because of his conservative views and prior votes as a court of appeals judge. Conservatives argued that evaluation should be limited to the nominee’s competence—that his ideology and prior votes should play no role in the Senate’s confirmation decision. Chemerinsky, supra note 23, at 624.
III. EFFECT ON JUDICIAL BEHAVIOR

In addition to addressing the effect of a judicial ranking system on the judicial selection process, Choi and Gulati raise important issues regarding the effect of a ranking system on judges’ behavior. There is substantial recent evidence that federal judges respond to incentives, including the opportunity to be appointed to a higher court. Therefore, the judicial incentives created by a ranking system must be considered.

A ranking system that rewards the publishing of opinions may cause some judges to publish more opinions. However, it is probably not desirable for judges to spend their (and their clerks’) time writing publishable opinions for legally insignificant cases. Such an incentive problem might be reduced by also including as a criterion the average number of citations to the judge’s opinions, because insignificant opinions are unlikely to be cited. However, including the average number of citations might encourage judges to use novel reasoning or reach atypical conclusions so as to increase the number of citations to their opinions. Choi and Gulati discount the possibility that judges will react this way. They argue that novel opinions are less likely to be cited by other judges. This is likely true in absolute terms; fewer judges are likely to agree with an unusual opinion. However, an opinion that takes a common approach to an issue will have much more competition. Judges can choose to cite either that opinion or, instead, to cite one of many other opinions making the same point. By definition, however, an unusual approach is one that has not been adopted by many courts, and thus a judge wishing to cite to such an opinion will have fewer to choose among. Because fewer such opinions exist, they may be more likely to be cited than an opinion that adopts a widely shared view.

For the same reason, even if no judges respond to a ranking system by publishing more novel opinions, including the number of cita-

28. S. Scott Gaille, Publishing by United States Court of Appeals Judges: Before and After the Bork Hearings, 26 J. LEGAL STUD. 371 (1997) (finding that federal appellate judges reduced their publication of books and articles after the confirmation hearings for Judge Bork’s Supreme Court nomination focused negatively upon parts of his publications); Gregory C. Sisk et al., Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning, 73 N.Y.U. L. REV. 1377, 1465-70 (1998) (finding that federal district judges with a higher chance of promotion to a U.S. court of appeals were more likely to find the politically popular Federal Sentencing Guidelines constitutional); Ahmed E. Taha, Publish or Paris?: Evidence of How Judges Allocate Their Time, 6 AM. L. & ECON. REV. 1, 15, 22 (2004) (finding that federal district judges with a higher chance of promotion were more likely to publish their opinions regarding the Federal Sentencing Guidelines’ constitutionality).

29. Choi and Gulati suggest that using the number of citations as a criterion may be a solution to the similar problem of judges writing numerous low-quality opinions quickly to boost their publication rate. Choi & Gulati, supra note 1, at 311-12.

30. Id. at 311 n.30.
tions as a criterion might result in creative judges receiving higher merit rankings. Although creative thinking is desirable in many fields, it may not be desirable in the judiciary. In fact, “exceptionably able” judges are viewed suspiciously by their peers, who suspect that such judges have an agenda. This suspicion may be unwarranted, but novel opinions are not necessarily better opinions. However, a ranking system that rewards frequency of citation may implicitly make that assumption.

Other criteria that likely would be included in the ranking also could create unintended and undesirable incentives. For example, Choi and Gulati suggest as a criterion the percentage of the judge’s decisions that are reversed by the circuit en banc or by the Supreme Court. Reversal rate may provide useful information regarding how frequently a judge commits errors. However, using it as a ranking factor also provides an incentive for a judge to make decisions that the majority of the other judges in the circuit and the current majority of the Supreme Court will agree with, rather than what the judge thinks are the correct decisions. In addition, using reversal rate will unfairly disadvantage a judge who has a political or judicial philosophy different from that of the majority of the rest of the circuit and of the Supreme Court.

Choi and Gulati acknowledge this problem and thus suggest giving more weight to reversals by judges with the same political orientation as the judge. For example, a Democratic judge who is reversed by a court with a majority of Democratic judges would have her ranking reduced more than if she were reversed by a court with a majority of Republican judges. Although Choi and Gulati do not suggest it, one could imagine a similar adjustment for the number of citations to a judge’s opinions: unfavorable citations of a Democratic judge by other Democratic judges would adversely affect that judge’s ranking more than would unfavorable citations by Republican judges.

However, such adjustments might actually increase the political partisanship of the judiciary. For example, if error is assumed when a judge disagrees with other judges of the same party, then judges will feel pressure to decide cases consistently with how other judges of the same party do. For example, a particular Republican judge may generally make politically conservative decisions, except on criminal procedure cases. A ranking system that penalizes him more for reversals of his decisions by other Republican judges than by De-

32. Choi & Gulati, supra note 1, at 307.
33. See id. at 310 n.29.
34. The phrase “Democratic judge” refers to a judge appointed by a Democratic President, and “Republican judge” refers to a judge appointed by a Republican President.
Democratic judges will create an incentive for him to also make politically conservative decisions on criminal procedure cases.

Choi and Gulati recognize the importance of judges' political independence. In fact, they suggest that the rankings should reward judges who rule differently from other judges who were appointed by a President of the same party.35 However, as just noted, this would also be inconsistent with giving more weight to reversals by judges of the same political party. This inconsistency illustrates another difficulty in constructing a valid ranking system: some objective measures of merit are subject to multiple, and even contradictory, interpretations.

Another reason to be concerned by a ranking system is its injection of competition into an appellate judiciary which relies upon cooperation and collegiality. Competition might undermine collegiality in a circuit. Many federal appellate judges have written about the importance of collegiality among judges on a court.36 As Judge Edwards points out, "In a collegial environment, divergent views are more likely to gain a full airing in the deliberative process—judges go back and forth in their deliberations over disputed and difficult issues until agreement is reached."37

The effect of collegiality can be seen in recent empirical scholarship on U.S. courts of appeals verifying that interaction with colleagues affects judges' decisions. In fact, in some areas of the law, the political orientation of the other two judges on a three-judge panel is at least as good a predictor of a judge's vote as is the judge's own political orientation.38 For example, a Republican judge sitting on a three-judge panel with two Democratic judges is more likely to vote

35. Choi & Gulati, supra note 1, at 310.
37. Edwards, supra note 36, at 1646.
to uphold an affirmative action program than is a Democratic judge sitting on a panel with two Republican judges.39

This collegiality could be undermined if judges respond to the incentives created by a tournament. For example, imagine that Judge A is on a three-judge panel with Judges B and C, and Judge C is assigned to write the majority opinion in a noteworthy case that is likely to be cited by other courts. If Judge A joins the majority opinion, she will receive no credit in the rankings for citations to that opinion, even if her comments were crucial in shaping Judge C’s opinion.40 However, if Judge A instead writes a concurring or dissenting opinion, then Judge A might be cited by other courts and have her ranking increased.

Choi and Gulati argue that because judges hear multiple cases together, such uncooperative behavior will be deterred, because other judges may retaliate by engaging in similar conduct when the uncooperative judge must write the majority opinion. Indeed, the fact that judges on the same court are repeat players with each other may deter some uncooperative behavior. However, this is likely more true in circuits that have relatively few judges. In these smaller circuits, the judges hear cases with the same judges more often than in larger circuits, and thus the effect of repeated interactions should be greater.41 As Judge Coffin states, “The difference in the collegial atmosphere between sitting with all of one’s colleagues each month and sitting with each only once or twice or even three times a year is enormous.”42

In addition, repeated interaction cannot be expected to prevent all judges from attempting to manipulate their rankings. Despite the benefits of working together collegially, even under the current system, some courts are more collegial than others. The incentives cre-

39. Sunstein et al., supra note 38, at 319; see also Revesz, supra note 38, at 1719 (finding that in environmental regulation cases before the U.S. Court of Appeals for the D.C. Circuit, “the party affiliation of the other judges on the panel has a greater bearing on a judge’s vote than his or her own affiliation”). Note that this also indicates that judges frequently cross party lines, a practice, as argued above, that could be deterred by a rating system.

40. COFFIN, ON APPEAL, supra note 36, at 221 (recalling that “on a significant number of occasions [upon receiving a draft opinion from a colleague], responding judges have been able to present a new way of looking at a case, or a hitherto overlooked case authority, or some undervalued fact or procedural point, and . . . a writing judge has gracefully changed course”).

41. Because cases in U.S. courts of appeals are generally heard by three-judge panels, a judge is typically sitting with two other judges. Because the composition of these panels rotates, judges in circuits with fewer judges will hear cases with a particular judge more often than if they were both in a circuit with more judges. For example, a judge sitting in a circuit with only six judges would sit on a panel with another particular judge approximately forty percent of the time. In a circuit with twelve judges, a judge would sit with another particular judge only approximately eighteen percent of the time.

42. COFFIN, ON APPEAL, supra note 36, at 216.
ated by a ranking system would create additional pressure on collegiality.

Finally, any incentive effects of a ranking system are limited by the fact that few judges have a realistic chance of being appointed to the Supreme Court. For many judges, the chance of nomination is so remote that it will not affect their behavior. However, Choi and Gulati point out that a ranking system might still provide positive incentives for those judges as well. They argue that those judges may be motivated to compete on these criteria simply out of concern that their ranking might affect their reputation.

Most judges care about their reputation, “both with other judges, especially ones on the same court . . . [and] with the legal profession at large.” However, this concern may not translate into judges being motivated by a ranking system, especially if judges and the broader legal profession do not believe that the ranking is an accurate measure of a judge’s quality.

Currently, a judge’s reputation is primarily based upon the impressions of those who work with the judge, litigate in front of the judge, or read the judge’s opinions. One must question whether these impressions of a judge’s quality would be significantly altered by an imperfect ranking system. This is particularly true if, when the rankings are made public, some highly regarded judges have lower rankings than lesser-regarded judges.

Choi and Gulati respond that, nevertheless, a ranking might motivate judges because law students who are potential law clerks may prefer to work for a higher-ranked judge. Law students generally lack familiarity with particular judges and thus are more likely than others to be influenced by the rankings. To secure the most desirable law clerks, a judge might be motivated to take actions that increase his ranking.

This raises the possibility that a ranking will become influential not because it is accurate but because it is the only information available to uninformed law students. We have seen this phenome-

43. Posner, supra note 31, at 111. Choi and Gulati note that the high value of the tournament’s prize—a spot on the Supreme Court—may still motivate some judges who have only a small chance of winning. Choi & Gulati, supra note 1, at 314-15. This may be true of some judges; however, many other judges in the middle of the rankings and below will be aware that they do not have any realistic chance of winning the tournament and thus likely will not be affected by the tournament.
44. Choi & Gulati, supra note 1, at 313-14.
47. Choi & Gulati, supra note 1, at 314.
non before: despite being heavily criticized as being inaccurate, the annual *U.S. News and World Report* law school rankings are important. Their importance results largely from the great attention they are given by prospective law students, who have little other basis for comparing law schools. These rankings lead some schools to take actions to unfairly manipulate their rankings. We should be reluctant to create a similar situation in the federal judiciary.

### IV. A Tournament of U.S. District Court Judges?

For many of the reasons discussed above, a stronger case exists for a tournament for positions on U.S. courts of appeals than for a tournament for positions on the U.S. Supreme Court. U.S. court of appeals judges often come from the ranks of U.S. district court judges, and a tournament of U.S. district court judges would lack many of the problems of a tournament of U.S. court of appeals judges. In addition, federal district judges should be more likely than U.S. court of appeals judges to respond to the incentives created by a tournament.

Many of the problems of measuring merit are less likely to affect a tournament of U.S. district court judges. Because federal district judges generally make decisions alone, their opinions are less likely to be affected by the preferences of their colleagues. In addition, they have almost complete discretion over choices such as whether to publish an opinion in a particular case. Thus quantitative measures of judicial performance will be measures only of a particular judge's behavior rather than also of the influence of other judges on a panel.

Also, unlike U.S. courts of appeals, the U.S. Supreme Court has almost complete control over its own docket. Thus, Supreme Court decisions are more likely to reflect policy judgments than are the decisions of courts of appeals. This means that reversals of federal dis-

---


50. For example, some schools have allegedly temporarily hired recent graduates to inflate the percentage of graduates who are employed, which is one component of the ranking. For other examples of how some law schools have attempted to manipulate their *U.S. News* rankings, see Dale Whitman, *Doing the Right Thing*, NEWSLETTER (ASS’n Am. Law Sch., Washington, D.C.), Apr. 2002, at 1, 1-4.


strict courts are less likely to reflect political disagreements than are reversals of federal appellate courts. Thus, the rate at which a judge is reversed would be a better—though still imperfect—measure of the merit of federal district judges than of federal appellate judges.

In addition, although collegiality can affect the operation of U.S. district courts, it is more important at the appellate level; U.S. district judges generally decide cases alone. Thus, concerns about the possible adverse effects of a tournament on collegiality are less important for U.S. district courts.

Finally, a federal district judge’s chance of promotion to a U.S. court of appeals is much higher than a federal appellate judge’s chance of promotion to the U.S. Supreme Court. There are 167 authorized judgeships in the twelve geographic circuits and only nine Supreme Court Justices, a ratio of almost 19 to 1. However, there are 676 authorized U.S. district court judgeships, which means that there are only about four district court judgeships for every U.S. court of appeals judgeship. This lower ratio means that federal district judges have a more realistic chance of promotion to a U.S. court of appeals than do federal appellate judges to the U.S. Supreme Court. Thus they would be more likely to respond to the incentives created by a ranking system.

V. CONCLUSION

Although the Supreme Court Justices should not be chosen by a tournament of judges, Choi and Gulati may have started a valuable process. Their proposal has focused attention on defining the characteristics of a good judge and on how to measure those characteristics. Although objective measures of merit will always be limited and imperfect, I share Choi and Gulati’s optimism that even flawed measures of merit will encourage the development of better measures. As these measures improve, they may provide valuable information to the judicial selection process. One can envision these measures eventually supplementing, though not replacing, the subjective ratings given to judicial nominees by the American Bar Association’s Standing Committee on the Federal Judiciary. The development of a well-

54. Because federal district judges are typically promoted to the U.S. court of appeals that covers the district they are in and court of appeals judgeships are normally reserved for a judge from a particular state, the true ratio facing a particular district judge will differ. Siak et al., supra note 28, at 1424.
55. Of course, if U.S. court of appeals judges value a position on the U.S. Supreme Court more than federal district judges value a position on a U.S. court of appeals, then the effect of this higher probability of winning the tournament will be at least partially offset.
56. The ABA evaluates a judicial nominee’s integrity, professional competence, and judicial temperament by reading the nominee’s legal writings, conducting exten-
regarded, objective ranking system might pressure Presidents to explain why they nominated low-ranked candidates and pressure Senators to explain why they oppose highly ranked candidates. As a result, the judicial nomination and confirmation process might eventually put more emphasis on nonpolitical factors.
KEEPPING SCORE: THE UTILITY OF EMPIRICAL MEASUREMENTS IN JUDICIAL SELECTION

DAVID C. VLADECK*

I. INTRODUCTION ................................................................................................. 1415
II. THE JUDICIAL SELECTION QUAGMIRE .......................................................... 1418
III. PARTISANSHIP'S IMPACT ON THE COURTS ....................................................... 1422
IV. CAN EMPPIRICL MEASURES COMBAT JUDICIAL PARTISANSHIP AND ENHANCE QUALITY? .............................................................................................................. 1427
   A. Quality of the Judicial Product ........................................................................ 1431
      1. Citations ................................................................................................. 1432
      2. Reversal Rates ....................................................................................... 1434
   B. Caseload Performance .................................................................................. 1435
   C. Independence ............................................................................................. 1437
V. CONCLUDING THOUGHTS ................................................................................. 1441

I. INTRODUCTION

The debate played out in the pages of this Symposium is a critical and timely one. The fight over judicial selection, especially over nominees to the federal courts of appeals and the U.S. Supreme Court, has become so politically polarized that, even during a time of one-party rule, it is a quagmire that many nominees fail to navigate successfully. My limited contribution to this Symposium is to offer the perspective of a full-time litigator who, for quite parochial reasons, is deeply invested in the quality of the federal appellate bench. In my view and in the view of most of the litigators with whom I have discussed this matter, the current approach to judicial selection—which places preeminent importance on a candidate’s ability to pass certain political litmus tests—has and will continue to erode the quality of the federal bench and the public’s confidence in the courts.

What politicians and academics often lose sight of is that litmus-test issues which make or break judicial appointments, particularly of appellate judges, determine the outcome of only a small proportion of the cases that reach the courts of appeals. Admittedly, these litmus-test issues comprise a higher percentage (but by no means all) of the cases before the Supreme Court. Competence, not political orien-

* Associate Professor of Law and Co-Director, Institute for Public Representation, Georgetown University Law Center. Prior to joining the Georgetown faculty, I was a staff attorney at, and then director of, Public Citizen Litigation Group, a nonprofit public interest law firm in Washington, D.C. During my twenty-six year tenure there, I argued several cases in the United States Supreme Court, nearly fifty cases in the courts of appeals, and eight cases before state courts of last resort. I also supervised a heavy volume of appellate litigation. The comments in this Essay are drawn from my litigation experience and from extensive discussions on judicial selection issues I have had with lawyers engaged in civil litigation in law firms, corporate law departments, the Department of Justice, and the public interest bar. I make no pretense of speaking on their behalf; but I believe that the sentiments expressed in this Essay are widely shared within the litigation bar.
tation, is what litigators and their clients most want from judges. This view is shared even by litigators who handle highly visible and often politically charged cases. Spare us the inept, plodding, or even mediocre judges appointed because they can be trusted to be the faithful guardians of whatever political issue is the cause du jour (fill in the blank: the fate of Roe v. Wade,\(^1\) affirmative action, the death penalty, same-sex marriage, states’ rights, judicial “activism,” and so on). Give us instead smart judges who will make an effort to put ideology aside and decide cases on the merits. After all, most lawyers believe (or have persuaded themselves to believe) that smart, fair-minded judges will decide the case in their favor. They do not believe they need the edge that ideologically sympathetic judges might give them, and they are wary of judges who seem to be ideologically out of step with the position they are asserting. Partisanship has no legitimate place in the courtroom.

This Symposium explores whether the identification of, and systematic and overt reliance on, a set of objective criteria would help those engaged in the appointment process\(^2\) better evaluate sitting federal court judges for elevation to higher judicial office. Studies have shown that Presidents are increasingly turning to the federal court bench to fill court of appeals and Supreme Court vacancies,\(^3\) in part because the nominees’ judicial records make predicting their fidelity to crucial political norms easier and more accurate. The subtext, of course, is that focusing on empirical measures of competence rather than single-issue political tests might help break the ideology-above-all-else logjam that has paralyzed the appointment process.\(^4\) I have no illusion that the President or the Senate will ever put politics aside in judicial appointments; but shifting the locus of the de-

---

2. Those involved in the appointment process include the President, his staff, members of the Senate, their staffs, the staff of the Senate Judiciary Committee, and interest groups concerned with judicial selection.
4. This is not an accusation directed at only one side of the fight. For instance, President George W. Bush has said that Roe v. Wade was wrongly decided, that he opposes abortion, and that he wants to appoint judges who agree with that sentiment. It is hardly surprising, therefore, that the Democrats on the Senate Judiciary Committee routinely try to pin down nominees on Roe. For example, during the confirmation hearings on Miguel Estrada, Senator Dianne Feinstein asked Mr. Estrada a series of questions about Roe, culminating with, “Do you believe that Roe was correctly decided?” Jeffrey Toobin, *Advice and Dissent: The Fight over the President’s Judicial Nominations*, NEW YORKER, May 26, 2003, at 42, 46. When Mr. Estrada provided what the Democrats viewed to be an unsatisfactory and vague answer, they demanded that the White House turn over memos Mr. Estrada had written while serving in the Solicitor General’s office that addressed Roe and ultimately blocked his nomination by sustaining a filibuster. *Id.* at 46-47. Both sides, of course, blamed the other for politicizing the appointment process.
bate, even at the margins, to nonpolitical issues would be a welcome step forward.

Having said that, I remain skeptical that all of the attributes we want to see in a judge (intellect, industry, integrity, impartiality, open-mindedness, judgment, compassion, and, at least for appellate judges, collegiality) are “objective” in the sense that they may be measured and quantified. For this reason, I see the suggestion that we engage in what Professors Stephen Choi and Mitu Gulati describe as a “Tournament of Judges”? in Swiftian, not literal, terms. Choi and Gulati acknowledge they have no expectation that the President, Congress, and the States will agree to amend the Constitution to select Supreme Court Justices through the use of an algorithm rather than the President’s Article II appointment power.6

Even if they did seriously advocate such an approach, the criteria that Choi and Gulati would employ in constructing their formula seem ill-suited to the job of predicting top-notch judicial performance, especially for Supreme Court nominees. On this point, my views are closer to those of my Georgetown colleague Professor Steven Goldberg, who argues that the talents one needs to be a superior Supreme Court Justice are light years apart from those needed to stand out on the courts of appeals, and thus even a stellar performance on the court of appeals is no predictor of success on the high court.7 Nonetheless, I agree with the fundamental insight of Choi and Gulati’s theory, namely, that there are real possibilities for measuring the performance of appellate judges to shed light on their suitability for the Supreme Court and that these possibilities should be explored.


6. Indeed, they suggest that they simply “advance a normative claim” but “see nothing unconstitutional about the president voluntarily looking to a set of objective measures to back up his claim regarding a candidate’s merit” or “anything problematic with the Senate looking to such measures as part of the fulfillment of its role.” Id. at 303 n.11.

7. Steven Goldberg, Federal Judges and the Heisman Trophy, 32 FLA. ST. U. L. REV. 1237 (2005). While I agree with Goldberg’s thesis, I think that the illustration he has chosen to make this point—the relative lack of success of Heisman Trophy winners in professional football—is ill-suited for his purpose. As Goldberg acknowledges, the Heisman Trophy is not awarded based on objective performance measures of the sort that Choi and Gulati champion (for example, yards gained for a running back, sacks for a defensive end) but is instead a beauty contest in which nearly a thousand sportswriters vote on the basis of whatever criteria each sportswriter deems relevant. See id. at 1243 n.27. Indeed, given the lackluster performance of Heisman winners in professional ranks, perhaps the Downtown Athletic Club of New York, which awards the Heisman Trophy, should take a page from Choi and Gulati and adopt a selection system that relies on quantitative measures of excellence rather than the subjective judgments of sportswriters.
carefully if for no other reason than that the alternative—perpetual partisan bickering and gridlock—is so unappealing.  

There also is an intuitive appeal to Choi and Gulati’s claim that at least some aspects of judicial quality can be measured. After all, other institutions—from large corporations to professional baseball teams—use empirical measurement to make hiring and advancement decisions. Other countries base their judicial systems on a professional corps of judges who are selected, rewarded, and promoted based, in part, on empirical measures. And, for that matter, those involved in the enterprise of selecting judges in the United States have long used empirical data of one sort or another to evaluate candidates for judicial office. There is no a priori reason why we should not consider adding to those measures if possible, but as I make clear below, I have real doubts that Choi and Gulati’s approach is ready even for a test run.

II. THE JUDICIAL SELECTION QUAGMIRE

Political strife over judicial nominees is nothing new. According to Professor Stephen Carter, “Inquiry into the nominee’s substantive legal positions became the order of the day in the wake of Dred Scott v. Sandford when all nominees were for a time called upon to prove their antislavery convictions.” Recent appointment decisions have been overtly political, even, at times, at the expense of competence. None of this should be surprising. Nearly ninety percent of

8. This is not an idealistic or Utopian conception of the appointment process. Even if it were possible for the President and the Senate to simply block out ideological considerations, and it is not, I do not contend that ideology is an inappropriate factor for the President to consider in selecting nominees or for the Senate to consider in performing its advice and consent function. Ideology is inevitable, and ideology matters. A person’s ideology may well influence how that person will vote on important matters. And it is therefore entirely appropriate for the President and the Senate to evaluate the likely impact an appointment will have. See Erwin Chemerinsky, Ideology and the Selection of Federal Judges, 36 U.C. Davis L. Rev. 619 (2003). What I object to, and what Choi and Gulati decry, is the obsessive and often exclusive focus not on a candidate’s ideology or overall judicial philosophy but on one or a few discrete battleground political issues, often to the exclusion of the candidate’s qualifications for office.

9. See infra notes 65-66 and accompanying text.

10. See infra notes 67-70 and accompanying text.


12. One engaging account of the Supreme Court selection process comes from John Dean’s book on President Nixon’s appointments to the Supreme Court. John W. Dean, The Rehnquist Choice: The Untold Story of the Nixon Appointment That Redefined the Supreme Court (2001). Dean’s book is an unusually rich historical source because not only was Dean personally involved in the selection process, but he also had ac-
the Justices appointed to the Supreme Court have been members of the President’s own political party, and there is copious evidence attesting to the importance of ideology when Presidents recruit and screen potential appointees and when the Senate considers nominations.13 There is every reason to believe that the battle over judicial nominations will only become more intensely partisan,14 with sev-

cess to all of the Nixon papers and audiotapes relating to the appointments. Id. at xv. Dean focuses on five weeks in the fall of 1971, between the time that Justices Hugo Black and John Harlan announced their retirements and President Nixon selected Lewis Powell and William Rehnquist to replace them. Among the candidates Nixon seriously considered were Virginia Congressman Richard Poff, a well-respected, conservative member of the House Judiciary Committee, who had virtually no experience practicing law, id. at 34, 39; Hershel Friday, a lawyer from Little Rock, Arkansas, who handled only bond work and had no litigation experience, id. at 166-68; and Mildred Lillie, who was an experienced trial and appellate court judge from California, id. at 175-77. None of these candidates were pushed because of his or her qualifications for the job; and no one would have argued that they were the best candidates to serve on the Supreme Court. They were serious contend-
ers because of their political orientation; their connection to senior administration officials; and for Poff and Friday, their status as Southerners, and for Lillie, her gender. Id. Tennes-
see Senator Howard Baker, who at the time was new to Washington, D.C., was offered a seat on the Court, even though Nixon knew only that he was moderately conservative and had some legal experience. Id. at 203-07. His status as a respected Southerner and his relative youth were seen as key attributes justifying his appointment. Id. Dean’s book makes it clear that while Nixon wanted competent appointees, he cared far more about their political credentials than anything else. Perhaps Rehnquist made this point most elo-

quently. During the White House’s search for candidates, a reporter asked him whether he was a possible nominee. Rehnquist said no: “I’m not from the South, I’m not a woman, and I’m not mediocre.” Id. at 191 (internal quotation marks omitted); see also Alan B. Morrison, The Rehnquist Choice, 55 STAN. L. REV. 1457, 1465 (2003) (book review).


14. One way to measure the intensity of the partisanship over judicial selection is to examine the rhetoric used by opponents of candidates nominated for what their opponents believe are overtly political reasons. For example, the People for the American Way, a lib-
eral group, warns members that:

Right-wing radicals have made it clear what the filibuster fight is really about—it’s about more than a few federal nominees—it’s about changing the rules when the rules don’t suit, it’s about the desire for absolute control, and it’s about what the Right Wing sees as the ultimate prize, the Supreme Court.

Over the past few months, the right-wing rhetoric has escalated: from the power-hungry assault on the filibuster to the more recent attacks on judges and judicial independence itself.

See Independent Judiciary, The Judiciary and the Right Wing: Attacks on Independence, People for the American Way, at http://www.pfaw.org/pfaw/general/default.aspx?oid=18471 (last visited June 1, 2005). Another liberal organization that works on judicial selection isues, the Alliance for Justice, has issued a similar action:

‘The president’s intention to renominate these highly controversial candidates ensures that the confirmation process will remain divisive,’ said [Alliance for Justice Director] Aron. ‘One would think the Bush administration would want to bring the country together after such a contentious reelection’ . . . ‘If the president tries to pack the courts with conservative extremists beholden to special interests and committed to turning back the clock on Americans’ rights, fair-minded senators must invoke their constitutional “advise and consent” re-

ponsibility and stand firm against such nominees,’ added Aron.

eral Supreme Court vacancies likely to come open in the next few years.15

Compounding the problem, partisan confirmation battles—once generally reserved for Supreme Court nominees—are now the norm for nominees to the lower federal courts.16 This, too, should be no surprise. Historically, Presidents have appointed lower court judges from their own political parties,17 and at least since the first Reagan Administration, Presidents have selected nominees based on publicly announced ideological criteria—President Reagan sought to appoint conservative judges who were not judicial activists, while President Clinton wanted judges who would be “strong supporters of Roe v. Wade.”18

http://www.independentjudiciary.org/news/release.cfm?ReleaseID=148 (last visited June 1, 2005). In fairness to these organizations, when President Clinton was making appointments, the rhetoric of conservative groups opposed to his nominees, like the Committee for Justice, was every bit as harsh, and their criticism of liberal groups is, by any measure, strident:

This radical coalition and its Senate allies, which cut its teeth attacking Supreme Court nominees Robert Bork and Clarence Thomas, has declared war on President George W. Bush’s nominees to the federal courts, especially to the important Circuit Courts of Appeal. While they cannot block every nominee, they have convinced Senate Judiciary Committee Democrats to attack and delay many of the brightest and most talented nominees, particularly those who are young minorities and women, to prevent them from becoming future potential Supreme Court candidates.

The Judicial Confirmation Battle, The Committee for Justice, at http://committeeforjustice.org/contents/confirmations (last visited Apr. 2, 2005). My point here is not to take sides in this battle or to criticize the work of these organizations, but it is simply to point out that the major combatants in the judicial selection battle see the fight in overtly political terms.

15. Chief Justice Rehnquist’s battle with cancer has only fueled speculation that a vacancy on the Court will arise as early as the summer of 2005. Accompanying the news of the Chief Justice’s illness has been a rising debate within the Senate about the propriety of using a filibuster to block a floor vote on judicial nominees. Many in the Senate leadership have suggested that changes should be made to the Senate’s rules to forbid the use of a filibuster on judicial appointments, although any change to longstanding Senate rules is likely to engender deep partisan divisions, so much so that the proposal is now referred to as the “nuclear option.” See, e.g., Charles Babington, Frist Likely to Push for Ban on Filibusters; Failure Risks Conservatives’ Ire; Success May Prompt Legislative Stalemate, WASH. POST, April 15, 2005, at A4; David D. Kirkpatrick, Lobbying Heats Up on Filibuster Rule Change, N.Y. TIMES, Apr. 3, 2005, at A24; Patricia A. MacLean, Bush’s Nominees Likely to Put Stamp on Circuits; GOP Judges May Go From 60% to 85%, NAT’L L.J., Mar. 28, 2005, at 1.

16. See, e.g., Toobin, supra note 4, at 42 (reporting on the Senate Judiciary Committee Hearing on the nomination of Leon Holmes for a district court judgeship in Arkansas and noting the strong objection of Committee Democrats because Holmes had publicly opposed the availability of abortion for rape victims on the ground that “conceptions from rape occur with the same frequency as snow in Miami”).

17. See Brudney, supra note 13, at 154 & nn.13-14.

18. Id. at 155 n.18 (emphasis added) (internal quotation marks omitted). President Carter formed the United States Circuit Judge Nominating Commission in 1977, which was an effort to place an emphasis on qualifications and diversity in the judicial selection process. The Commission had a bipartisan membership and did not apply avowedly political criteria in identifying candidates for court of appeals appointments, although it was
The heightened partisanship over lower court appointments has weakened the traditional system for nominating lower court judges—which placed principal reliance on selection by home state Senators.\textsuperscript{19} I do not suggest that Senators selected judicial candidates without reference to ideology. But there were countervailing factors. Not only were some Senators politically more moderate than the President, but also the selection process generally involved consultation between the home state Senators in an effort to reach an accommodation. After all, because one home state Senator could always “blue slip” (automatically block) a colleague’s appointment, there were substantial incentives to reach an agreement on nominees. These factors placed a check on purely ideological appointments.\textsuperscript{20} While home state Senators still have considerable influence in selecting lower court nominees, the White House now spearheads these selections to ensure that the President’s nominees are on board with him ideologically.\textsuperscript{21}
III. PARTISANSHIP’S IMPACT ON THE COURTS

Unrestrained partisanship in the judicial selection process has had a number of adverse effects on the judiciary, apart from creating gridlock and long-unfilled judicial vacancies, which take their own toll. For one thing, the intensely partisan nature of today’s appointment process is a serious disincentive for even the most highly qualified individuals to throw their hat into the judicial-appointment ring. Not only does the process threaten to be emotionally bruising, but many nominees also have to place their careers on hold for a year or more while the process grinds ahead, only to be rejected for political reasons or to see their nominations lapse because Congress goes out of session or the President’s term expires.

Out-and-out partisanship also breeds friction in the courts, which, in turn, breeds disrespect for the law. Consider the fractious history of the D.C. Circuit in the five years from 1986 to 1991. During that time, the political balance on the court was in flux and ultimately shifted from a majority of Democrat-appointed judges to one of Republican-appointed judges. Relations among the judges became strained and reached their nadir when Judge Laurence H. Silberman, during a post-argument conference, told Judge Abner Mikva that “if you were ten years younger I would be tempted to punch you in the nose.” According to press accounts, Judge Silberman for reasons of ideological orientation than qualification—a conclusion not inconsistent with that of Professor Epstein and her colleagues.

Finally, even conceding that the Senate generally is sensitive to the qualifications of Supreme Court nominees, that conclusion hardly supports the contention that the process is not a “mess.” The partisanship that infects the appointment process cannot be parsed that finely; there is no separate appointment process for Supreme Court nominees. To be sure, if one looks just at outcomes for Supreme Court appointments—are the men and women confirmed highly qualified?—the system appears to work tolerably well (putting aside, as do the authors, whether one believes that competence rather than excellence is the right standard). But if one examines the political cost of the appointment system for all judicial nominees, the descriptive “mess” is a euphemism, not an overstatement.


23. Id.

24. Ann Pelham, Silberman, Dogged by Story, Provides Details of Outburst, Legal Times, Mar. 11, 1991, at 7 [hereinafter Pelham, Silberman Provides Details of Outburst] (internal quotation marks omitted) (printing a letter sent to The New York Times by Judge Silberman about the incident). Relations between these two judges remained strained and once again flared up while Judge Mikva was serving as Chief Judge. After a leak of information about a draft affirmative action opinion written by then-Judge Clarence Thomas, Judge Silberman called for a formal investigation—a request Chief Judge Mikva rejected. Judge Silberman then released a public request for a formal hearing, which criticized Judge Mikva’s handling of the matter. The request was signed by the other Republican appointees on the court. See Ann Pelham, Gloves Off as Judges Get Personal, Legal Times, Feb. 24, 1992, at 7. The feud continued, even to the point that eight years later, Judge Silberman and the other Republican appointees on the court boycotted the D.C. Circuit’s dedication ceremony of Judge Mikva’s official portrait. See Jonathan Groner, Partisan Gap at Mikva Ceremony, Legal Times, Oct. 23, 2000, at 8.
berman was not kidding. The bickering within the court was so heated that it did not remain behind closed doors. The judges used increasingly acerbic language in their opinions to criticize the views of their colleagues. The drastic increase of en banc hearings and calls for rehearings en banc reflected the shifting political orientation of the court, further impaired the court’s collegiality, and


27. See, e.g., Stuart Taylor Jr., Ideological Feud Erupts in a Key Appeals Court, N.Y. TIMES, Aug. 15, 1987, at A7 (pointing out caustic nature of the judges’ opinions); Associated Press, Court Upholds Tougher Fuel Ratings, N.Y. TIMES, May 18, 1988, at A18 (noting that the opinions “[d]isplay[ed] a bitter ideological split”).


brought about wholesale changes in the law on issues both important and trivial.\textsuperscript{29}

Unfortunately, the acrimony that marked this period for the D.C. Circuit was not an isolated occurrence. Recently, the Sixth Circuit experienced its own bout of internecine warfare. In an en banc decision in \textit{Grutter v. Bollinger},\textsuperscript{30} the court upheld the University of Michigan Law School’s use of affirmative action policies in admission, but two conservative members of the court dissented. Judges Danny J. Boggs and Alice M. Batchelder, both Republican appointees, dissented in part to accuse Boyce F. Martin, Jr., the Chief Judge of the Court and a Democratic appointee, of jury-rigging the composition of the court that would consider the case en banc. Judges Boggs and Batchelder alleged that Chief Judge Martin delayed the circulation of en banc petitions in violation of Sixth Circuit procedures until two conservative judges, who might have tipped the balance on the court in favor of the dissenters, took senior status and were therefore ineligible to participate in en banc proceedings. Judge Boggs described what occurred as “very significant and obvious violations of rights of members of [the] court.”\textsuperscript{32}

Chief Judge Martin did not respond to these charges. But Judge Karen Nelson Moore did, calling the accusation “shameful”\textsuperscript{34} and noting that it “marks a new low point in the history of the Sixth Circuit,” which “will irreparably damage the already strained working relationships among the judges of [the] court.”\textsuperscript{35} Judge Eric L. Clay also responded to the accusation, chiding the dissent for having “chosen to stoop to such desperate and unfounded allegations”\textsuperscript{36} that “unjustifiably distort” what transpired. This was not the first time that partisan battles had poisoned the atmosphere in the circuit.\textsuperscript{38}

\textsuperscript{29} Many of the court’s en banc decisions wrought major changes in the law. See, e.g., Critical Mass Energy Project v. Nuclear Regulatory Comm’n, 975 F.2d 871 (D.C. Cir. 1992) (en banc) (changing the standards that govern public availability of business information submitted to the government). But some of the court’s en banc cases involved questions that were remarkably trivial, such as whether nonprevailing plaintiffs in actions under the Freedom of Information Act, 5 U.S.C. § 552 (2000), were liable for taxable costs under Rule 54 of the Federal Rules of Civil Procedure, which at most amount to $50 to $100. Baez v. United States Dep’t of Justice, 684 F.2d 999 (D.C. Cir. 1982) (en banc).

\textsuperscript{30} 288 F.3d 732 (6th Cir. 2002) (en banc), aff’d, 539 U.S. 306 (2003).

\textsuperscript{31} \textit{Id.} at 810-14 (Boggs, J., dissenting); \textit{see also id.} at 815 (Batchelder, J., dissenting).

\textsuperscript{32} \textit{Id.} at 810-14 (Boggs, J., dissenting); \textit{id.} at 815 (Batchelder, J., dissenting).

\textsuperscript{33} \textit{Id.} at 815 n.49 (Boggs, J., dissenting).

\textsuperscript{34} \textit{Id.} at 753 (Moore, J., concurring).

\textsuperscript{35} \textit{Id.} at 758.

\textsuperscript{36} \textit{Id.} at 772 (Clay, J., concurring).

\textsuperscript{37} \textit{Id.} at 758.

\textsuperscript{38} \textit{See, e.g.}, Memphis Planned Parenthood, Inc. v. Sundquist, 184 F.3d 600 (6th Cir. 1999) (order denying rehearing en banc); \textit{id.} at 607 (Boggs, J., separate statement) (arguing that colleagues’ remarks directed at him are “unwarranted and unprofessional”); \textit{id.} at 608-09 (Batchelder, J., separate statement) (admonishing Judge Keith for resorting to “an
Partisanship also affects outcomes, which I suppose is entirely the point, but in ways that unjustifiably destabilize the law. I start from the premise (perhaps naïvely) that members of the federal judiciary strive, by and large, and generally with success, “to decide cases in accord with the law rather than with their own ideological or partisan preferences.” To be sure, there are exceptions (generally the cases I lose). Most litigators believe, however, that panels composed entirely of judges appointed by one President are more likely to vote on ideological grounds than mixed panels. Professor Erwin Chemerinsky makes the point well: “When I talk to a lawyer who is about to have an argument before a federal court of appeals, the first question I always ask is: who is your panel? That is because ideology matters so much in determining the result in so many cases.” Emerging empirical research suggests that political dominance on any court encourages judges to act on the basis of ideology rather than principle, especially when there is no countervailing influence on a panel and little or no likelihood of further review, by either the full court or the Supreme Court. Professor Richard Revesz demonstrated that the D.C. Circuit’s approach to environmental cases during the mid-1980s

ad hominem attack on specifically named members of [the] court); id. at 604 (Keith, J., dissenting) (criticizing the “seemingly careless fashion in which the majority recklessly disregards the constitutional rights afforded to minor girls, a group of people who are the least able to defend their rights”). The recent en banc practice in the Sixth Circuit has been the subject of a number of academic studies. See, e.g., Pierre H. Bergeron, En Banc Practice in the Sixth Circuit: An Empirical Study, 1990-2000, 68 TENN. L. REV. 771 (2001).

39. Harry T. Edwards, The Judicial Function and the Elusive Goal of Principled Decisionmaking, 1991 WIS. L. REV. 837, 838. Not surprisingly, judges insist that this is the case. See id. at 838 (acknowledging nonetheless that it is not “entirely clear . . . that partisan politics and ideological maneuvering have no meaningful influence on judicial decisionmaking”); see also Patricia M. Wald, Regulation at Risk: Are Courts Part of the Solution or Most of the Problem?, 67 S. CAL. L. REV. 621, 645 (1994). Many academics, however, believe that any suggestion that judges can put ideology aside in ruling on cases is a fairy tale:

The argument is that ideology has to be hidden from the process to limit the likelihood that once on the bench judges will base their decisions on ideology. This argument is based on numerous unsupportable assumptions: it assumes that it is possible for judges to decide cases apart from their views and ideology; it assumes that judges do not already often decide cases because of their views and ideology; it assumes that considering ideology in the selection process will increase this in deciding cases. All of these are simply false. Long ago, the Legal Realists exploded the myth of formalistic value-neutral judging.

Chemerinsky, supra note 8, at 630-31.

40. Chemerinsky, supra note 8, at 628.

through early-1990s depended heavily on the composition of the panel hearing the case, with panels having two or three Republican appointees far more likely to invalidate EPA action than panels composed of two or three Democratic appointees.42

Professor Cass Sunstein’s more recent research similarly shows that “a panel of . . . like-minded judges tends to go to extremes.”43 Sunstein dubs this process “ideological amplification” and contends that “[i]n many areas, an individual Republican appointee sitting with two other Republican appointees is far more likely to vote in the stereotypically conservative fashion than an individual Republican appointee who sits with one Republican and one Democrat.”44 Sunstein summarized his research to Senate staffers as follows:

Republican nominees really are different from Democratic nominees. And when you get three Republican nominees sitting together . . . they get really conservative. We had a large enough sample, so that we could see proof of this in affirmative action, campaign finance, the Americans with Disabilities Act, judicial review of environmental regulations, and sex discrimination.45

Sunstein acknowledges that the amplification effect applies to Democratic appointees as well.46 He argues, and I agree, “that a high degree of diversity on the federal judiciary is desirable, that the Senate is entitled to promote reasonable diversity, and that without such diversity judicial panels will inevitably go in unjustified directions.”47 Revesz and Sunstein’s work, important as it is, is descriptive; they leave unaddressed the question of how to accomplish these goals, which Choi and Gulati take as their starting point.

42. Revesz, Environmental Regulation, supra note 41, at 1741-44. Unlike Sunstein, Revesz does not find the presence of a Democratic appointee on a panel with two Republican appointees to be a significant moderating factor. Id. at 1761-63 & n.83.
44. Id. at 167.
45. Toobin, supra note 4, at 42 (internal quotation marks omitted); see also Sunstein, supra note 43, at 168-76.
47. Id. at 190. My experience is consistent with the evidence amassed by Revesz and Sunstein, although I am inclined to agree with Sunstein’s observation that mixed panels tend to be more cautious than panels composed entirely of one party’s appointees. But I think that both Revesz and Sunstein miss one point that perhaps defies measurement: judges are not fungible and cannot be counted on to act in accordance with the ideology of the President who appointed them (which is why President Eisenhower reportedly said that appointing Chief Justice Earl Warren was his “biggest mistake” and that appointing Justice Brennan was his “second biggest mistake”). My intuition is that judges appointed mainly because of their partisan views maintain that approach on the bench and more reliably vote in a partisan manner, while judges selected mainly for their merit, but are viewed as ideology safe, tend to put ideology aside far more often. See Choi & Gulati, supra note 5, at 318 n.48; Kenneth L. Manning et al., George W. Bush’s Potential Supreme Court Nominees: What Impact Might They Have?, 85 Judicature 278 (2002). My submission, of course, is that we should find a way to tilt the judicial selection process to favor the latter category of judges and weed out the former.
IV. CAN EMPIRICAL MEASURES COMBAT JUDICIAL PARTISANSHIP AND ENHANCE QUALITY?

By now it should be clear that I believe that the judicial selection process is tilted too heavily toward political factors. I recognize that there are nominees who both share the President’s views on important political issues and are exceptionally well qualified for the federal bench. But that is far from universally true. Too many judges are selected not because they are excellent candidates and the President is certain that they will vote the “right” way on the key issues, but because they are perceived to be loyal protectors of the President’s ideological goals. Even if I am wrong and the roster of judges appointed for reliability reasons alone is short, it is still too long.

Consider one example. The most devastating indictment of a judicial appointee I have seen was published by an unlikely source: The New Yorker. Its “Talk of the Town” column critiqued President George H.W. Bush’s final appointment to the Second Circuit, Dennis Jacobs. After calling Judge Jacobs’ appointment a “curious choice,” the column said that Judge Jacobs “has none of the qualifications that the public and, certainly, most lawyers assume are required for a federal judgeship as important as one on the Second Circuit.” Judge Jacobs had spent his entire career with a large New York law firm engaged in “the narrow field of reinsurance litigation,” where the cases, “while important to the companies involved, are of little importance to the public.” The column also pointed out that

[he] was never a judge on a lower court, a criminal prosecutor, or a defense attorney. He never worked for the government or per-
formed any kind of public service. He never participated in a civil-rights case or in any case where constitutional issues played a major role. And, unlike many federal judges, he has shown no interest in legal scholarship.52

As The New Yorker put it, Judge Jacobs’ submissions to the Senate Judiciary Committee “provide a picture of a prosperous lawyer of limited interests and little distinction.”53 The “single clue to why he was nominated,” The New Yorker added, was that he identified the Federalist Society as the only organization to which he belonged. The New Yorker noted that President Reagan had earlier appointed a number of distinguished judges to the Second Circuit, including Yale Law School Professor Ralph Winter and the well-regarded “law-and-order trial judge[,] . . . George C. Pratt.”54 Bush’s appointment of Jacobs, said the magazine, showed that “the ranks of conservatives who were both eminent and reliable had thinned” and that “President Bush was ultimately forced to settle for reliability alone.”55

To me, settling for reliability alone is a nonstarter. Judging today is a challenging enterprise. Long gone are the days (if they ever existed) when complex statutory cases were the exception and not the rule, when oral argument stretched on for hours, and when there was time for leisurely reflection. Even apart from the crushing caseloads that make reflection a luxury, a substantial percentage of the cases coming before federal judges are enormously complicated. Judges are routinely called upon to unravel the mysteries of the complete pre-emption doctrine in ERISA cases56 and resolve questions as esoteric as whether an expired statute nonetheless had the force of law.57 Judges must also confront the shifting tides in Eleventh Amendment jurisprudence,58 decide when a statute enacted by Congress applies retroactively, even when Congress itself gave the question not a mo-

52. Id.
53. Id.
54. Id. at 31-32.
55. Id. at 32.
57. See Wis. Project on Nuclear Arms Control v. United States Dep’t of Commerce, 317 F.3d 275 (D.C. Cir. 2003); see also id. at 285 (Randolph, J., dissenting) (observing that, according to the majority, “[t]he statute has expired but its legislative history is good law”). Having served as co-counsel for the losing party in the case, my view of course is that the case was wrongly decided.
ment of thought, and deal with a perennial source of judicial irritation—the Sentencing Guidelines—which, even in this post-Booker environment will be the subject of continuing judicial attention.

Judging today calls for bright, tenacious, energetic lawyers who can work collegially and commit themselves to the endeavor of resolving difficult problems in a reasonable period of time. These are attributes that are easy to label, hard to define, and harder still to measure. The problem with the debate on judicial selection is that it often proceeds along these lines—quality is elusive, subjective, and therefore defies measurement. Another variation is that, although perhaps the quality of lawyering can be measured to some extent, it is far from clear that the qualities that make a good lawyer make a good judge, or a capable court of appeals judge worthy of a seat on the Supreme Court.

Choi and Gulati see these as defeatist arguments. They concede that there are many attributes judges should have that are subjective, but they contend that there are many objective factors that reflect the quality of judges that can be quantified and measured. Choi and Gulati are not alone in this conviction. Some industrialized democracies do not draw judges from the ranks of lawyers but instead select and train an independent judiciary. Germany, for example, has

---


61. Several of my Georgetown colleagues have challenged the assertion that litigators want “smart” judges. They argue that rational litigators want judges ideologically in sync with the arguments they assert, which they believe—and Revesz and Sunstein’s research supports—maximizes their chance of success. Perhaps if all the lawyers cared about was winning one case, there might be force to this argument. But that is rarely so. Most litigators concentrate in certain areas, often for repeat clients. It is myopic to believe that these lawyers and their clients always, or even generally, care about one case to the exclusion of those that will follow. Poorly reasoned decisions may be great fodder for academics, but they are stumbling blocks for litigators who seek to shape the law. Indeed, the only thing worse than losing a case for the wrong reason is winning a case for the wrong reason.

Moreover, this argument overlooks that ideological dominance is a two-way street. I would prefer smart judges to overtly political judges in part because I am far more likely to see appellate panels composed of a majority of judges appointed by Republicans than Democrats, who my colleagues would forecast would be more sympathetic to the cases I litigate. But given the ideological swings that mark the courts, most litigators believe that they would be better off with stability and predictability, rather than the dramatic shifts in the law that occur when courts undergo an ideological transition. Indeed, nothing is quite as fleeting as success bestowed by an ideological panel; as soon as the ideological pendulum swings in the other direction, these gains are quickly transformed into losses. See, e.g., Critical Mass Energy Project v. Nuclear Regulatory Comm’n, 975 F.2d 871 (D.C. Cir. 1992) (en banc) (splitting on political lines and overturning longstanding precedent).
a professional corps of judges who are selected from recent law school graduates on the basis of an exam: “[O]nly the graduates with the best examination results have any chance of entering the judicial corps.”62 German judges are promoted on the basis of an “efficiency rating” [that] is based in part upon objective factors, such as caseload discharge rates and reversal rates, and in part on subjective peer evaluation.”63 The judicial selection and promotion process in Japan appears to be similar.64 Large law firms, corporations, and other institutions use objective measures in making hiring and promotion decisions.65 Even managers of professional baseball teams have turned to objective criteria in player selection; this approach is epitomized by Billy Beane’s striking success in raising the lowly Oakland Athletics to playoff contenders by relying on hard data on player performance rather than the intuition and judgment of scouts in drafting and trading for players.66

Those engaged in judicial selection have also used objective measures to evaluate judges, although not in the systematic way proposed by Choi and Gulati. For sitting district court judges, the Administrative Office of the U.S. Courts maintains extensive case-specific statistics about each judge’s caseload, disposition rate, the time motions have been pending, and so forth—information that speaks volumes about a judge’s productivity.67 Similar statistics are compiled on the performance of circuit court judges.68 Other statistics available through LEXIS and Westlaw show the number of opinions a judge

63. Id. at 850 (footnote omitted).
66. See MICHAEL LEWIS, MONEYBALL: THE ART OF WINNING AN UNFAIR GAME (2003); see also Note, Losing Sight of Hindsight: The Unrealized Traditionalism of Law and Sabermetrics, 117 HARV. L. REV. 1703 (2004); Supreme State: What if Supreme Court Justices Were Picked Based on Their Career Numbers Instead of Their Politics?, LEGAL AFFAIRS, Sept./Oct. 2004, at 32; Benjamin Weiser, Judge’s Decisions Are Conspicuously Late, N.Y. TIMES, Dec. 6, 2004, at A1 (using statistics from the Administrative Office of the U.S. Courts to show that one district judge sitting in the Southern District of New York had 289 motions in civil cases pending for more than six months, the highest in the nation by far, and more than three times the number of any of the judge’s colleagues on the Southern District).
67. Although the statistics are available on the Internet in aggregate form only, they are compiled and available to judges, and presumably Senate Judiciary Committee staffers, for each judge. See, e.g., Federal Court Management Statistics 2004: District Courts, at http://www.uscourts.gov/cgi-bin/cmsd2004.pl (last visited Apr. 7, 2005).
writes, how that number compares to the productivity of other judges on the same court, how often the judge is affirmed and reversed, how often a judge’s opinions are cited by his colleagues, and so forth. The Senate Judiciary Committee has long used a detailed form to collect basic information about nominees that includes a wide range of objective information—including all of the information that The New Yorker thought was lacking with the elder Bush’s final appointee to the Second Circuit.69 So there is at least some readily available objective data that sheds light on a judge’s qualifications for higher office.70

Choi and Gulati identify three criteria that would also tell us a good deal about possible Supreme Court nominees, although they hedge by saying that “[c]rafting a fully specified system of objective factors that may be used to rate judges . . . is beyond the scope of [their article].”71 The criteria they identify for preliminary inquiry are the following: (a) the quality of the judicial product; (b) caseload performance; and (c) independence.72 In the abstract, of course, these are all important factors that bear on a judge’s suitability for elevation. Nonetheless, I have reservations about Choi and Gulati’s selection of these factors for appellate judges who act collaboratively, and I believe that substantial refinements would be needed to make each of these criteria useful measures for appellate judges.

A. Quality of the Judicial Product

In defining what they mean by the “quality of the judicial product,” Choi and Gulati point to two main factors: (a) the extent to which a judge’s opinions are cited, not just in judicial opinions, but in academic writings and even casebooks (with adjustments made for favorable and unfavorable treatment of the opinion); and (b) a judge’s reversal rate (with recognition that such a measure “may unfairly penalize judges with different political views from those on the Court”).73 I am skeptical that either of these factors will tell us as

69. See supra notes 49-55 and accompanying text.
70. Make no mistake, data of this sort are also used for overtly partisan purposes. Much of the debate over Judge Bork’s nomination to the Supreme Court was based on “objective” data about his extensive voting record while a judge on the D.C. Circuit. Compare THE WHITE HOUSE REPORT: INFORMATION ON JUDGE BORK’S QUALIFICATIONS, JUDICIAL RECORD & RELATED SUBJECTS (1987), and OFFICE OF PUB. AFFAIRS, U.S. DEP’T OF JUSTICE, A RESPONSE TO THE CRITICS OF JUDGE ROBERT H. BORK (1987), with COMM. ON THE JUDICIARY, U.S. SENATE, RESPONSE PREPARED TO WHITE HOUSE ANALYSIS OF JUDGE BORK’S RECORD (1987), and PUB. CITIZEN LITIG. GROUP, THE JUDICIAL RECORD OF JUDGE ROBERT H. BORK (1987). An abridged version of the White House report and full versions of the other three reports (with the exception of their respective appendices) are all reprinted in 9 CARDOZO L. REV. 185-508 (1987).
71. Choi & Gulati, supra note 5, at 305.
72. Id. at 305-13.
73. Id. at 306-07.
much about judicial quality for court of appeals judges as Choi and Gulati theorize.

1. Citations

Let us start with the basics—not all citations are created equally. For the most part, my sense about citations is summed up by Judge Leventhal’s famous line about resort to legislative history: it is “the equivalent of entering a crowded cocktail party and looking over the heads of the guests for one’s friends.”74 For routine citations—standards of review and settled legal propositions—most judges simply cite the most recent circuit precedent or their own most recent opinion. It is hard to ascribe any significance to citations of that sort. Where a citation is needed for a more controversial point, judges tend to cite the opinions of judges who they know and trust. But familiarity is hardly a proxy for quality.

Moreover, cases are often cited not because of the trenchancy or novelty of their judicial analysis but because of either their memorable language or their ability to distill a complicated legal issue down to a multifactor test that judges and lawyers embrace for the sake of convenience. For instance, without disputing for a moment Judge Learned Hand’s greatness, I would suggest that his opinions are cited at least as often for the nimbleness of his pen as they are for the force of his logic.75 Consider Judge Hand’s famous line in *Cabell v. Markham*:—an otherwise unremarkable case:

> It is true that the words used, even in their literal sense, are the primary, and ordinarily the most reliable, source of interpreting the meaning of any writing: be it a statute, a contract, or anything else. But it is one of the surest indexes of a mature and developed jurisprudence not to make a fortress out of the dictionary; but to remember that statutes always have some purpose or object to ac-

---

74. Conroy v. Aniskoff, 507 U.S. 511, 519 (1993) (Scalia, J., concurring in the judgment) (remarking on Leventhal’s use of this metaphor in regard to legislative history); see also Patricia M. Wald, *Some Observations on the Use of Legislative History in the 1981 Supreme Court Term*, 68 IOWA L. REV. 195, 214 & n.143 (1983). Although this description is universally attributed to Judge Leventhal and appears in countless opinions, albeit with minor variations, it does not appear that Judge Leventhal himself ever used it in a judicial opinion.

75. Judge Posner has suggested that one measure of Judge Hand’s greatness was his “gift of verbal facility that enables a familiar proposition to be expressed memorably, arrestingy, thus enforcing attention, facilitating comprehension, and, often, stimulating new thought (in which case the expressive dimension of judicial greatness merges with the creative).” Richard A. Posner, *The Learned Hand Biography and the Question of Judicial Greatness*, 104 YALE L.J. 511, 524 (1994) (book review). It may be that one measure of a Judge’s greatness is his or her verbal dexterity, but surely elegant writing, without more, is not the makings of a great judge.

76. 148 F.2d 737 (2d Cir. 1945), aff’d, 326 U.S. 404 (1945).
complish, whose sympathetic and imaginative discovery is the surest guide to their meaning.77

This passage was hardly a pathbreaking moment in the law. Other judges had made the same point, although less eloquently than Judge Hand. Nonetheless, according to LEXIS’s Shepard’s service, *Cabell v. Markham* has been cited no fewer than 331 times, including ten times by the Supreme Court. It may be that superior writing skills are indeed a fair measure of judicial performance, as Judge Posner claims in his defense of Hand’s greatness.78 But if that is Choi and Gulati’s thesis, they ought to say so.

There are also citations that are repeated time and again simply because they mark the distillation of existing circuit law into a formula that can easily be used by litigants and courts. For example, every circuit has one key case that reduces the requirements of a preliminary injunction into a familiar four-part test. In the D.C. Circuit, it is Judge Leventhal’s decision in *Washington Metropolitan Area Transit Commission v. Holiday Tours, Inc.*,79 which, according to LEXIS’s Shepard’s service, has been cited over 1200 times. Again, while I admired Judge Leventhal, I do not believe that his authorship of *Holiday Tours*—and especially the fact that it is cited frequently—demonstrates his excellence as a judge. *Holiday Tours* simply refined slightly the analysis the court had laid out years earlier in *Virginia Petroleum Jobbers Ass’n v. FPC*,80 another widely cited decision.81

Choi and Gulati are plainly correct that there are many opinions that are heavily cited because they break new legal ground. Perhaps they are right that these opinions are true measures of judicial quality and that it is fair to give the credit to the opinion’s author, and not the other judges who served on the same panel. But nothing in Choi and Gulati’s proposal would help us differentiate those opinions, which might speak volumes about the talents of the author, from other equally widely cited cases that do not.

A citation-dependent inquiry as to judicial quality would also not account for significant variations in the caseloads among the circuits, which is a problem that would need to be addressed. Choi and Gulati assume for statistical purposes that the circuits have comparable cases. But that is not so. The D.C. Circuit, for example, bears the brunt of record-heavy administrative law cases because the District

77. Id. at 739.
78. See Posner, supra note 75.
79. 559 F.2d 841 (D.C. Cir. 1977).
80. 259 F.2d 921 (D.C. Cir. 1958).
of Columbia is the seat of government, because the court is a convenient forum for those affected by regulation, and because a number of statutes assign exclusive jurisdiction to the D.C. Circuit. This leads to two consequences; one is that judges on the D.C. Circuit hear fewer cases than their counterparts elsewhere (about half the number of cases heard by judges on the Seventh and Eighth Circuits), and the other is that their opinions, which often address specialized issues, are less likely to be cited outside of the D.C. Circuit. Other circuits—like the Sixth—also fall outside the mainstream in terms of caseloads. Thus, unless adjustments are made for these variations, the judges on these circuits would be at a significant disadvantage if frequency of citation were a key measure of judicial quality.

Finally, measuring frequency of citation overlooks a core and perhaps lamentable fact about the way most courts of appeals do business. The overwhelming majority of cases today are resolved in unpublished opinions. Published opinions represent only about twenty percent of the courts' dispositions. Enough ink has already been spilled over the rule-of-law and fairness concerns related to unpublished decisions. But whatever one's views are on those issues, the presence of unpublished opinions has to be addressed by Choi and Gulati. If quality is to be measured by examining the work product of judges, we cannot simply ignore the bulk of their work product because it is unpublished and hard to compile. There ought not to be two tiers of justice in the United States. A judge who garners high marks for his or her published opinions should be held to an equally high standard for unpublished opinions. But, unless Choi and Gulati refine their approach, looking just to published opinions would give judges a free pass with respect to their unpublished rulings, only adding to existing incentives for judges to shortchange opinions that will not be published.

2. Reversal Rates

I am also skeptical that reversal rates would tell us much about the quality of court of appeals judges, but not solely for the reasons that Choi and Gulati mention. My concern is that most cases are reviewed by the Supreme Court only when a "mature" circuit split arises. Rarely does the Court grant review when just one circuit

---

84. Id.
85. They recognize that court of appeals opinions are collaborative products, and thus it might be unfair to attribute the ruling only to its author. And they acknowledge that such a measure "may unfairly penalize judges with different political views from those on the [Supreme] Court." Choi & Gulati, supra note 5, at 307.
parts company from its sister circuits. The Court generally lets these “immature” circuit splits simmer in the hope that the circuit courts will ultimately resolve their disagreement and come to a uniform position. By the time the Supreme Court grants review, there are generally several circuits that have taken sides on the issue and often there are several decisions within each circuit applying circuit law. At some point, of course, the Supreme Court grants one case to review. But it is more a matter of serendipity (or, if Choi and Gulati have their way, bad luck) than legal error for a judge to author a court of appeals decision that gets reversed by the Supreme Court. He or she is no more culpable than any of the other, and there may be many, court of appeals judges who have written opinions saying precisely the same thing. If reversal rates mean anything, then we need to apportion the “blame” more equitably and charge each circuit judge who backed the losing side with whatever demerit we assign.86

Nor am I persuaded that the Supreme Court always has the better argument. As Justice Jackson famously noted, “We are not final because we are infallible, but we are infallible only because we are final.”87 Indeed, there are times that even a Supreme Court opinion is not the last word on a legal question. Congress occasionally overrules the Court on statutory matters,88 and the Court, at times, reverses itself on constitutional issues.89 In those cases, would Choi and Gulati give the court of appeals judge who is ultimately vindicated extra credit?

B. Caseload Performance

Choi and Gulati next point to factors that might reveal the effort a judge puts into his or her job. For a Supreme Court Justice, Choi and Gulati theorize that a Justice exerting “maximal effort” would use “clerks minimally, dissent as often as would be warranted, and vote

86. Consider just one example. By the time the Supreme Court decided Cipollone v. Liggett Group, Inc., 505 U.S. 504 (1992), at least six federal circuit courts had concluded that the Federal Cigarette Labeling and Advertising Act preempted all state law tort claims brought by cigarette smokers, id. at 508-09 nn.2 & 3. No circuit court took a different view, although a number of state courts had ruled that the Act did not preempt tort suits. Id. Is it fair to make Judge Edward Becker, the highly distinguished judge who authored the Third Circuit’s opinion that was reversed in part in Cipollone, shoulder all of the blame, especially since the Third Circuit applied what was then fairly settled law on the preemption issue?


to grant certiorari in as many cases as is possible.”90 They argue that to get an insight on how a court of appeals judge might perform on the Supreme Court, we should look at the number of opinions a judge authors, whether the judge or law clerks were the main authors, and the overall number of cases the judge played a part in deciding in a specified period of time.91 My problem with Choi and Gulati’s suggestion is not the measures of productivity they want to examine (although I am not sure I see the premium on judges rather than law clerks writing first drafts of opinions).92 Indeed, as I suggest earlier, I think that there are many measures of productivity that are worth exploring. After all, we want energetic and diligent judges, and there are many measurements available that bear on those attributes.

My concern is not with Choi and Gulati’s interest in measuring productivity but is instead with how they would define it, namely, that it is in the “public interest” for the Supreme Court to hear more cases than the eighty or so a year that currently make up its docket. The size of the Supreme Court’s caseload is an issue that has sparked enormous debate, especially as the Court has reduced its docket from approximately 150 cases per year to about 80 in the span of just a few years.93 With Justice White’s retirement, there is no longer a Justice on the Court who insists that it grant review every time a circuit split arises. While there are forceful arguments on both sides of the debate, I think that the burden rests with Choi and Gulati to defend the proposition that it would be, as they claim, “in the public interest” for the Court to “grant certiorari in as many cases as is possible.”94

90. Choi & Gulati, supra note 5, at 309.
91. Id. at 309-10.
92. I recognize that Choi and Gulati’s contribution to this Symposium is an extended explanation of why they believe that it matters whether judges draft their own opinions. I have doubts that it is necessarily the best use of a judge’s time to draft opinions word-for-word or to write first drafts of opinions. I am sure that many conscientious judges transform first drafts written by clerks into their own product through editing and revision. Writing is an idiosyncratic process, and I assume that judges find the right balance that works for them. Rather than trying to disentangle whether it is the judge or the law clerks responsible for opinion writing, it may be preferable to simply evaluate the quality of the final product. If a judge is not fully engaged in the opinion writing process, the final product is likely to suffer. I do, however, agree with Choi and Gulati that a judge’s disengagement from the opinion writing process is a warning sign that the judge is not acting diligently.
94. Choi & Gulati, supra note 5, at 309.
There are sound reasons that cut in the other direction. One is that a less interventionist Court is not necessarily a less effective or influential Court.\footnote{Justice Brandeis apparently remarked often that “[t]he most important thing we do is not doing.” See ANTHONY LEWIS, GIDEON’S TRUMPET 252 n.21 (Vintage Books ed. 1989) (1964) (explaining that Justice Brandeis’ former law clerk, Harvard Law Professor Paul A. Freund, recalled Brandeis frequently making this comment) (internal quotation marks omitted); see also Ashwander v. Tenn. Valley Auth., 297 U.S. 288, 346 (1936) (Brandeis, J., concurring).} Although this Court hears fewer cases than its immediate predecessors, it has honed its selectivity and expanded use of the GVR\footnote{GVR stands for grant, vacate, and remand, which is the order the Supreme Court enters when it sends a case back to a lower court for reconsideration in light of a recent ruling.} process to orchestrate sweeping changes in the law. It would be hard to contend that a reduced caseload has diminished the current Court’s oversight of the law. Another reason is that the Court has plainly made a conscious decision that it is better to allow circuit splits to become mature before intervening, thereby reducing the number of cases the Court reviews. Whether the Court does so to ensure full percolation of the legal issues before the Court intervenes, to see whether the problem resolves itself, or to avoid dealing with an issue that is too trivial to justify the Court’s time is a matter for debate. While one could disagree in principle with those judgments, they are surely worthy of some response by Choi and Gulati.

C. Independence

Choi and Gulati’s final criterion is judicial “independence.” They explain that “part of the judicial mission is to decide cases impartially”—“independent of political ideology.”\footnote{Choi & Gulati, supra note 5, at 310.} To determine independence, they propose to measure “the frequency with which the judge is in opposition to another judge selected by the same president (or a president from the same political party),” because “[j]udges who are systematically more willing to disagree with politically like-minded judges are . . . more unbiased in their approach to individual cases.”\footnote{Id.} I agree with Choi and Gulati that independence is a key attribute for judges and that examining dissents may tell us something about independence. But the test that Choi and Gulati propose is too crude to tell us much about a judge’s independence, and, as I explain, there are better measurements available.\footnote{Choi and Gulati’s use and definition of the word “independence” injects a degree of confusion. “Independence” in describing a judge can mean one of two things: a judge’s independence from the ideology of the President who appointed the judge; or independence from other judges appointed by the same President, or a President of the same political party, which is the meaning Choi and Gulati appear to ascribe to the word. In my view, for the word “independence” to be synonymous with “impartiality,” it must mean independence from the appointing President, a point I develop below.}
First, their test does not screen out cases that are apolitical, such as a breach of contract case between two large companies. Many cases decided by the courts of appeals carry no ideological freight, and thus the fact that judges appointed by the same President are on opposite sides of the question tells us little about the impartiality and independence of either judge. They just disagree. But their disagreement does not show a willingness to challenge orthodoxy or turn away from the political orientation of the President who appointed them.

Second, their test does not focus on cases in which judges depart from the ideological views of the Presidents who appointed them, let alone in those cases in which one might reasonably expect ideology to matter. Compounding the problem, even when the case does have clear ideological implications, the fact that the judges disagree does not mean—as Choi and Gulati suggest—that they should both be given credit for independence. It seems to me that the opposite is true; that is, the judge who departs from the position one would predict based on the ideological orientation of the appointing President is entitled to credit for independence, not the judge who toes the line regardless of whether he or she is in the majority or in dissent.

There are, to be fair, rare cases in which judges in both the majority and the dissent deserve credit for independence. Consider Wisconsin Project on Nuclear Arms Control v. United States Department of Commerce, a Freedom of Information Act (FOIA) case brought by an arms control group to get information from the Department of Commerce regarding the licensing of “dual-use” commodities (products that can be used for civilian and military purposes) for overseas sales. The government claimed the information was shielded from disclosure by virtue of an expired provision of the Export Administration Act that had been in effect only intermittently since 1977 and was not in effect at the time the information request was filed or the case litigated. The panel majority, Circuit Judge Judith Rogers (a Clinton appointee) and Senior Circuit Judge Stephen Williams (a Reagan appointee), concluded that, even though the Act had lapsed, it nonetheless barred disclosure of the records. Judge Raymond Randolph (a Bush appointee) dissented, chiding his colleagues: “The statute has expired but its legislative history is [still] good law. So say my colleagues, in a most curious opinion.”

100. Choi & Gulati, supra note 5, at 310 & n.29.
102. Id. at 278.
103. Id. at 278-79.
104. Id. at 281-85.
105. Id. at 285.
From Choi and Gulati’s standpoint, much more curious than the opinion is the lineup of the judges. Judge Rodgers authored the majority opinion and had been appointed by President Clinton, who was committed to openness under FOIA,\textsuperscript{106} while Judge Randolph had been appointed by President George H.W. Bush, who quickly reversed the openness policies of his predecessor.\textsuperscript{107} If one is measuring independence, then it would seem only fair that both Judge Rodgers and Judge Randolph get high marks for their views in this case, because both took positions that were at odds with the views of the Presidents who appointed them.

On the other hand, there are cases in which only one judge is, in fact, acting “independently,” albeit not as Choi and Gulati use that term. Consider the Fourth Circuit’s denial of the petition for en banc rehearing in \textit{Hamdi v. Rumsfeld}.\textsuperscript{108} Hamdi was seeking a rehearing of the Fourth Circuit’s prior decision rejecting this petition for a writ of habeas corpus.\textsuperscript{109} The panel opinion, written by Chief Judge J. Harvie Wilkinson (a Reagan appointee) and joined by Judges William Wilkins and William Traxler (appointed by Reagan and Clinton, respectively), upheld the government’s position that the detainees imprisoned at the Guantanamo Naval Base in Cuba were not entitled to a hearing and directed that Hamdi’s petition for a writ of habeas corpus be dismissed.\textsuperscript{110} Dissenting from the denial of Hamdi’s petition for an en banc rehearing, Judge J. Michael Luttig (a Bush appointee) argued that the government could not make its case for detention without giving Hamdi a chance to be heard.\textsuperscript{111} Judge Luttig’s position was directly at odds with that taken by a Republican President in a case all about ideology. While it would make sense to


\textsuperscript{107} Memorandum from John Ashcroft, Attorney General, U.S. Dep’t of Justice, to Heads of All Federal Departments and Agencies (Oct. 12, 2001), http://www.usdoj.gov/oip/foia/post/2001foia/post19.htm (rescinding Reno Memorandum and assuring agencies that the Department will defend agency withholding decisions “unless they lack a sound legal basis or present an unwarranted risk of adverse impact on the ability of other agencies to protect other important records”).

\textsuperscript{108} 337 F.3d 335 (4th Cir. 2003).


\textsuperscript{110} \textit{Id.}

\textsuperscript{111} \textit{Hamdi}, 337 F.3d at 357-68 (Luttig, J., dissenting from denial of rehearing en banc). Judge Diana Gribbon Motz, a Clinton appointee, also dissented. \textit{Id.} at 368-76. Because her dissent is presumably in keeping with the ideology of the President who appointed her, it is hard to see why she should get credit for independence, even though her dissent put her on the other side of the issue from Judge Robert King, the other Clinton appointee on the Court, who did not dissent from the court’s denial of rehearing en banc.
credit Judge Luttig for his independence (especially if he harbored hopes of ascending to the Supreme Court), the same certainly cannot be said for Chief Judge Wilkinson, whose opinions in the case adhered faithfully to the party line. Nonetheless, under Choi and Gulati’s test, they would be given equal credit for independence.

Finally, there are cases in which judges demonstrate independence even when they are in the majority and there are no dissents. Consider *Public Citizen v. Burke*, which was a politically charged case brought to challenge a Department of Justice Office of Legal Counsel (OLC) interpretation of regulations issued by the National Archives and Records Administration under the Presidential Recordings and Materials Preservation Act of 1974—the law that empowered the government to take possession of the records of the administration of Richard M. Nixon. The OLC had issued an opinion requiring on constitutional grounds that the regulation be interpreted as obliging the archivist to acquiesce in any claim of executive privilege asserted by the former President to block release of the materials—an opinion that would apply not just to Nixon’s records but to the records of any former President. The case was heard by Circuit Judges Laurence Silberman and David Sentelle (both Reagan appointees) and District Judge Harold Greene (appointed by President Carter), sitting by designation. In a unanimous opinion written by Judge Silberman, the court held the OLC memorandum could not justify deferral to privilege claims by former Presidents, finding the OLC position to be based on a “misunderstanding of the Constitution.” Once again, if one is measuring a judge’s willingness to take a stand on an important legal question that has evident ideological overtones, then one has to give credit at least to Judge Silberman for his opinion in *Burke*, and perhaps as well to Judge Sentelle, because he did not dissent but instead joined with Judge Silberman in opposing a measure strongly supported by President Reagan.

The upshot of this discussion underscores my initial point: Dissents can (but not always) tell us something about independence, but only a small part of the story. The better way to evaluate independence is to determine when, and how frequently, judges depart from the position that is ideologically aligned with that of the Presidents who appointed them in cases where both ideology and outcome matter. To me, that is the true measure of independence, especially for a judge being considered for higher judicial office.

---

112. 843 F.2d 1473 (D.C. Cir. 1988).
114. *Burke*, 843 F.2d at 1474-75.
V. CONCLUDING THOUGHTS

At bottom, my heart is with Choi and Gulati because I share their distaste for the current partisan logjam that we call the judicial selection process. It is brutal and unfair to nominees. It undermines collegiality. It breeds distrust of the legal system. And it is woefully inefficient. Judicial vacancies go unfilled for years or more, clogging dockets, crowding courts, and delaying justice. So I want Choi and Gulati to be right; there must be a better way to select judges. But my mind says not yet. They propose a “Tournament of Judges,” but at this point, they have no satisfactory way of keeping score. A tournament needs a clear-cut winner. To be sure, there are measures of judicial quality and productivity available for social scientists to review and perhaps use to rank judges on these attributes, as Professor Landes115 and others have done. But those measurements are nowhere as comprehensive or as accurate as would be needed to run the sort of tournament Choi and Gulati propose, let alone to displace the more judgment-laden selection process that is now employed. And in fairness to Choi and Gulati, they make no pretense otherwise.

Rather, they want to debate whether it is worthwhile to try to shift the emphasis from partisanship to quality in judicial selection and whether it is realistic to believe that the intensely political process now used to select judges will yield to a less partisan approach. I do not have an answer to either question. But there is, of course, a middle ground. There is no reason why efforts like Choi and Gulati’s to identify objective measures of judicial quality cannot be used now to enrich the often arid debate over judicial selection and to shift focus away from single-issue litmus tests. Choi and Gulati have shown that, with some fine-tuning, we might be able to examine factors like citations, reversal rates, opinions issued, opinions authored, and the like, and draw conclusions about judicial quality, productivity, and independence. Why not look to these measures when scrutinizing nominees? This approach would fall short of Choi and Gulati’s stated goal of a rank-order tournament from which one clear winner would emerge, but it might accomplish their real goal of focusing the judicial selection process on hard facts rather than harmful rhetoric.

And isn’t that the point. No one—not even Choi and Gulati—believes that it is possible to take the politics out of judicial appointments (even if we wanted to). The real question is what can be added to the debate to ensure that concerns about quality and competence are not sacrificed amidst the political infighting over nominees. “Demonstrated excellence, not merely promise, is what must be re-

---

115. See supra note 82.
quired for the most demanding legal job in the land.”¹¹⁶ Here Choi and Gulati’s insight that we must formulate and evenhandedly apply a consistent set of criteria for judicial selection is irresistible. And in many ways, the questions posed by The New Yorker to evaluate Judge Jacobs’ appointment seem suitable additions to those proposed by Choi and Gulati: has the nominee (1) served in the government as a judge or a prosecutor or served as a defense lawyer; (2) engaged in litigation on constitutional issues or questions of importance to the public; (3) participated in legal scholarship or other academic pursuits; (4) had a varied legal career; (5) performed community service; and (6) demonstrated the intellect, industry, impartiality, and judgment that we should require for judicial appointees.¹¹⁷ I do not suggest that the presence of these factors guarantees greatness. I do suggest that the absence of these factors is a danger signal that the President has selected the nominee for reasons of “reliability alone,” and that the Senate would be well served to challenge the nominee’s qualifications for service, not for political reasons but simply on the basis of merit.