

JUDGING THE JUDGES

FRANK B. CROSS†

STEFANIE LINDQUIST††

ABSTRACT

The evaluation of judges, especially circuit court judges, has commanded increased attention, with the quantitative analyses of Stephen Choi and Mitu Gulati. However, the proper dimensions for the evaluation of judges remains much disputed. Critics have challenged Choi & Gulati's scales for measuring judicial quality but have offered little that is positive that would improve measurement. The critics make philosophical challenges to whether the measures truly capture the qualities of judging we should desire, but they offer no measurement tools to improve on Choi and Gulati.

We hope to advance the theoretical and empirical evaluation by incorporating different scales for evaluating judges. We consider Choi & Gulati's data, plus the record of each judge's opinions on review at the Supreme Court and other more subjective measures of judicial quality. We then employ a cluster analysis to differentiate among different types of judges on all these dimensions. This analysis does not provide a rating of the "best" judges, because the standards for "best" judging are contested. Some judges may be considered best, because their opinions receive more citations. However, these very judges may be considered less than best, because they render unnecessarily expansive decisions that yield more citations. Some argue for judges with minimalist decisionmaking characteristics that tend to result in fewer citations. Our cluster analysis simply categorizes circuit court judges into groups with like decisionmaking

Copyright © 2009 by Frank B. Cross and Stefanie Lindquist.

† Herbert D. Kelleher Centennial Professor of Business Law, McCombs School of Business, University of Texas at Austin; Professor of Law, University of Texas Law School; Professor of Government, University of Texas at Austin.

†† Thomas W. Gregory Professor of Law, University of Texas Law School; Professor of Government, University of Texas at Austin.

characteristics. This enables an analysis of which group, or type of judge, should be considered “best.”

TABLE OF CONTENTS

Introduction	1384
I. Choi and Gulati’s Tournament and Other Evaluations.....	1387
A. Choi and Gulati’s Methods and Results.....	1387
B. Use of the Results for Supreme Court Appointments	1388
1. Particularized Methodology Critiques.....	1388
2. Broader Usage Critiques.....	1393
C. Other Research Assessing Judges.....	1400
II. Evidence from Supreme Court Review.....	1402
A. Significance of Supreme Court Review.....	1403
B. Data and Methods.....	1406
C. Results.....	1406
III. Judicial Types Compared and Contrasted.....	1414
A. Judicial Types	1415
1. Judicial Entrepreneurs	1419
2. Judicial Minimalists	1422
B. Cluster Analysis of Judges and Types	1425
IV. Important Additional Research.....	1429
Conclusion.....	1436

INTRODUCTION

The rigorous comparative evaluation of federal judges has become a popular topic in academic research; the topic’s popularity can be traced in large part to a series of articles written by Professors Stephen Choi and Mitu Gulati evaluating federal circuit court of appeals judges as candidates for Supreme Court appointment.¹ Their research has already generated a considerable amount of analytical research specific to their model, some of it critiquing their methods or conclusions.² Yet these critiques often overlook some important

1. 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, *Choosing the Next*]; Stephen J. Choi & G. Mitu Gulati, *Mr. Justice Posner? Unpacking the Statistics*, 61 N.Y.U. ANN. SURV. AM. L. 19 (2005) [hereinafter Choi & Gulati, *Mr. Justice Posner?*]; Stephen Choi & Mitu Gulati, Essay, *A Tournament of Judges?*, 92 CAL. L. REV. 299 (2004) [hereinafter Choi & Gulati, *A Tournament*].

2. See, e.g., Michael J. Gerhardt, *Merit vs. Ideology*, 26 CARDOZO L. REV. 353, 353–54 (2005); William P. Marshall, *Be Careful What You Wish for: The Problems with Using Empirical Rankings to Select Supreme Court Justices*, 78 S. CAL. L. REV. 119, 119–21 (2004); WERL, *On*

issues, and they have generally failed to add any meaningful descriptive data to inform the debate commenced by Choi and Gulati. This Article strives to advance the analysis by reanalyzing the issues and adding new empirical information for the evaluation of judges.

Like most of the other existing research, we focus on the federal circuit court judiciary. Professors Choi and Gulati chose to evaluate these judges because the federal appellate judiciary is the most likely source for Supreme Court appointments.³ In addition, the circuit court judiciary is probably the single most important level of the federal judiciary in light of its extensive caseload and policy making authority.⁴ Given the limited resources of Supreme Court review, the circuit courts “are largely left to themselves” to develop “legal rules in unsettled areas of law.”⁵ There is thus great value in understanding the relative quality of circuit court judges as these appointments, in the aggregate, are arguably more important for most federal court cases than any individual appointment to the U.S. Supreme Court.

In Part I, we review the background research on judicial quality. The Choi and Gulati study is our benchmark as it investigates the quality of federal appeals court judges in great detail and has received considerable attention and assessment. In Part I we also review and evaluate several critiques of this study’s methodology and usage. The methodological disputes raised in these critiques are central to any evaluation of the worth of this or any other quantitative measures of judicial quality. The Choi and Gulati study is not the only quantitative evaluation of circuit court judges, and therefore we also review several other approaches that have been taken to this project.

Tournaments for Appointing Great Justices to the U.S. Supreme Court, 78 S. CAL. L. REV. 157, 157–58 (2004); Lawrence B. Solum, *A Tournament of Virtue* 1–2 (U. San Diego Legal Studies, Research Paper No. 05-16, 2004), available at <http://ssrn.com/abstract=588322>. The Choi and Gulati tournament has spawned a symposium for an entire issue of a law review. Symposium, *Empirical Measures of Judicial Performance*, 32 FLA. ST. U. L. REV. 1001 (2005). Their research has also been discussed in the Chronicle of Higher Education. David Glenn, *Jousting with Gavels*, CHRON. HIGHER EDUC., May 23, 2003, at A16.

3. See Choi & Gulati, *Choosing the Next*, *supra* note 1, at 26 n.2.

4. See FRANK B. CROSS, *DECISION MAKING IN THE U.S. COURTS OF APPEALS* 1–2 (2007) (“[T]he circuit courts are much more important [than the Supreme Court] in setting and enforcing the law of the United States.”); Richard A. Posner, *Judicial Behavior and Performance: An Economic Approach*, 32 FLA. ST. U. L. REV. 1259, 1273 (2005) (“[T]he Supreme Court reviews only a minute percentage . . . of court of appeals decisions. Entire fields of law are left mainly to the courts of appeals to shape.”).

5. DAVID E. KLEIN, *MAKING LAW IN THE UNITED STATES COURTS OF APPEALS* 51 (2002).

Part II introduces new criteria for evaluating circuit court judges, criteria that were used in connection with the appointment of Samuel Alito to the Supreme Court.⁶ We examine the success of circuit court judges on Supreme Court review, quantifying how often the Court reversed or affirmed their opinions and decisions. This provides a metric that yields different results from those provided by Professors Choi and Gulati. The judges with the highest quality score for outside circuit citations (the measure pioneered by Choi and Gulati) did not tend to have greater success when the Supreme Court reviewed their decisions. Because Supreme Court success seems related to judicial quality for circuit court judges, this suggests that the earlier quality measures require more scrutiny before their accuracy may be presumed.

In Part III, we consider the different types of judges who serve on the circuit courts, using a variety of sources including our results on Supreme Court reversals, Choi and Gulati's quality measure, and other available measures reflecting circuit judge prestige and quality. Prior research shows that judges' role orientations differ. In Part III, therefore, we differentiate between judicial entrepreneurs and judicial minimalists and discuss how prior evaluations of judicial quality may actually capture a judicial role orientation rather than some element associated with intrinsic quality. We address the effect of judicial role orientation through a use of cluster analysis, which is a statistical tool for grouping observations with shared characteristics. In this way, we can identify categories of judges who may share similar approaches to the judicial role.

Part IV suggests directions for further important research on circuit court judges. A measure of judicial quality or judicial type can be very useful to future analyses of judicial decisionmaking. Such a measure can help define the characteristics of judicial decisionmaking by identifying judges who are especially ideological or deferential. The results can then be used to assist decisionmakers appoint judges of the sort society desires.

6. See Adam Liptak & Jonathan D. Glater, *Alito's Dissents Show Deference to Lower Courts*, N.Y. TIMES, Nov. 3, 2005, at A1 (referencing this study's findings on Supreme Court treatment of Alito's circuit court decisions).

I. CHOI AND GULATI'S TOURNAMENT AND OTHER EVALUATIONS

Various researchers over the years have undertaken to measure judicial caliber empirically, but Professors Stephen Choi and Mitu Gulati have produced the most prominent and important measure, which they initially styled as involving a “tournament” of judges.⁷ In this tournament, they created criteria to evaluate circuit court judges for promotion and applied those criteria to extensive data they collected on contemporary circuit court judges. They produced ordered results identifying the best judges on several scales.

A. *Choi and Gulati's Methods and Results*

Professors Choi and Gulati designed a disciplined system to evaluate circuit court judges for purposes of future Supreme Court appointments by first identifying factors considered relevant to judicial quality, then gathering data to measure those elements. They then adjusted their data to compare the scales and select the top judges.

Choi and Gulati measured the circuit court judges for “productivity, quality, and independence.”⁸ The productivity standard measures the number of published opinions the circuit court judge wrote during the time period in absolute number and adjusted to account for differing caseload and publication norms in each circuit court. The quality measure seeks to identify the judges who write the highest-quality opinions by measuring the relative number of citations to those opinions by other judges. In particular, Choi and Gulati focused on citations outside the judge’s own circuit to avoid the biases of specific intracircuit effects, on citations to the judge’s top opinions, and on invocations of the judge’s name in other opinions. As for independence, the authors measured the frequency with which judges disagreed with appointees of the same political party. This standard measures in significant part the judge’s willingness to separately concur or dissent, but it also measures the likelihood that a judge will dissent from (or concur with) an appointee of the same party, which demonstrates ideological as well as decisional independence.

The authors base their scales on circuit court opinions written between 1998 and 2000 for each of their scales and found, as a general result, that Judges Posner and Easterbrook of the United States

7. See Choi & Gulati, *A Tournament*, *supra* note 1.

8. Choi & Gulati, *Choosing the Next*, *supra* note 1, at 33.

Court of Appeals for the Seventh Circuit dominated the measures. Other judges, however, had higher independence scores.⁹ Nevertheless, when the scores were weighted to account for different criteria, Posner and Easterbrook became quite dominant.¹⁰

B. Use of the Results for Supreme Court Appointments

Professors Choi and Gulati's initial article on the subject proposed a tournament "where the reward to the winner is elevation to the Supreme Court."¹¹ Although the title of their subsequent article suggests that their results could be used to choose Supreme Court Justices, they concede that this is unlikely to occur and that it, indeed, was not their ultimate objective.¹² Instead, they propose that the results could at least influence the choice of Justices through their persuasive effect and potential political use.¹³ Nevertheless, various reviewers have criticized the Choi and Gulati research, at least as applied to the selection of Justices. Regardless of this debate, we believe it may have greater value in other uses, especially in the context of academic research on judicial behavior.

The Choi and Gulati study prompted a flurry of follow-up analyses. Numerous articles were written on the authors' methodology and the proposed use of their results. Although much of the commentary on the Choi and Gulati tournament have been quite positive, the commentators have made a number of very specific criticisms. These criticisms fall into two main categories: disputes about the accuracy of their measures of judicial characteristics (methodology critiques) and disputes about whether those measures are suitable for the selection of Supreme Court Justices (usage critiques).

1. *Particularized Methodology Critiques.* One set of criticisms of the tournament involves the methodology used to assign "virtue" to circuit court judges. The productivity, quality, and independence scales used by Professors Choi and Gulati are all relatively objective

9. *Id.* at 68–69.

10. *Id.* at 73.

11. Choi & Gulati, *A Tournament*, *supra* note 1, at 299.

12. Choi & Gulati, *Choosing the Next*, *supra* note 1, at 38.

13. Professors Choi and Gulati suggest that their standards "enable transparency in the nomination process" and create a "de facto presumption—one that the president has to rebut (or else face public pressure to the extent that the tournament's objective winners are easy to observe)." *Id.* at 81.

assessments, but according to critics, they incorporate subjective evaluations. Some have questioned whether the three criteria truly capture praiseworthy characteristics in the judiciary. Others have accepted the value of the criteria but question whether the authors' quantitative measures truly capture the relevant concepts.

Although the judicial productivity measure seems quite straightforward, Professor Daniel Farber questions whether the number of published opinions measure fairly captures judicial productivity.¹⁴ The more serious issue, however, is whether Professors Choi and Gulati's productivity measure is an effective means to identify qualified Supreme Court appointees, even if it is an accurate measure of productivity. The average circuit court judge in Choi and Gulati's sample wrote over thirty-two published opinions per year.¹⁵ Supreme Court Justices now write fewer than ten opinions per year.¹⁶ It hardly seems relevant to prioritize circuit court productivity as a standard for Supreme Court appointments when even the least-productive circuit court judges match all the Justices of the Supreme Court in opinion-writing frequency. Indeed, some circuit court judges may write fewer opinions because they devote greater time and care to crafting each individual opinion, a feature that should be considered a positive for appointment to the Supreme Court. The productivity measure may have uses, but it hardly seems salient for Supreme Court appointment evaluations.

One might also argue that the Choi and Gulati independence measure is a questionable measure of judicial quality. It rewards separate opinion writing, which is at best a debatable measure of quality, since collegiality and deference to other panel members may be a better measure. The partisan composition of the circuit courts may also skew these independence calculations. For example, if Republican judges predominate in the circuit courts, Democratic judges will systematically have lower independence scores, simply because they have relatively less opportunity to demonstrate independence from other Democratic appointees on the same panel.

14. Daniel A. Farber, *Supreme Court Selection and Measures of Past Judicial Performance*, 32 FLA. ST. U. L. REV. 1175, 1177 (2005). He suggests, for example, that the decision to publish precedents may reflect only the "self-centeredness" of the judge. *Id.* A count of unpublished opinions might be added to this measure.

15. Choi & Gulati, *Choosing the Next*, *supra* note 1, at 44.

16. Even if the Supreme Court were to double the number of cases it accepts, which Professors Choi and Gulati believe might be beneficial, its annual opinion output would still be much less than that of the median circuit court judge.

Professors Choi and Gulati apparently conceive of the independence measure as a test of non-ideological decisionmaking, which may be based on findings that appointees of the same party appear to influence one another and vote together.¹⁷ Yet the independence scale fails to measure this form of effect if there is a range of ideological preferences within the cohort of appointees of one party.¹⁸ Indeed, the measure would actually “reward” the most ideologically extreme of party appointees.¹⁹ Suppose that a particular circuit judge appointed by a Democratic president was extremely liberal. This judge might often dissent from the opinions of other Democratic appointees because their decisions were not liberal enough, thus revealing ideological extremity rather than any “true” ideological independence.

Moreover, the vast majority of circuit court decisions are not ideologically driven.²⁰ Consequently, a purely ideological judge would produce a very high independence score given that most of the decisions rendered by other appointees of the same party are far less ideological.²¹ A much simpler and more readily available test for non-ideological decisionmaking would directly analyze the ideological direction of the judge’s decisions and the extent to which ideology appears to explain those rulings.²² Thus the independence measure

17. See Frank B. Cross & Emerson H. Tiller, Essay, *Judicial Partisanship and Obedience to Legal Doctrine: Whistleblowing on the Federal Courts of Appeals*, 107 YALE L.J. 2155, 2168–73 (1998) (showing the effect of the ideological composition of court panels on judicial voting); Richard L. Revesz, *Environmental Regulation, Ideology, and the D.C. Circuit*, 83 VA. L. REV. 1717, 1751–58 (1997) (reporting the effects of other panel members on judicial decisions).

18. This association is apparent from casual observation of the Supreme Court. Justices Stevens and Souter were appointed by Republican presidents, as were Justices Scalia and Thomas, but the two sets frequently disagree with one another on ideologically patterned grounds. There are now “judicial common space” scores that place circuit court judges on a continuum of ideology, based on their votes, and there are significant differences even among appointees from the same president. Lee Epstein et al., *The Judicial Common Space*, 23 J.L. ECON. & ORG. 303, 304, 320 (2007).

19. See Marshall, *supra* note 2, at 127 (noting that the measure “rewards a judge who is outside the ideological mainstream”).

20. See Frank B. Cross, *Decisionmaking in the U.S. Circuit Courts of Appeals*, 91 CAL. L. REV. 1457, 1504–09 (2003) (demonstrating that ideology explains only a small portion of circuit judges’ voting behavior and has lesser explanatory power than legal factors).

21. There is some evidence that the Choi & Gulati scores are indeed influenced by this effect. See WERL, *supra* note 2, at 177 (noting that the highest independence scores were associated with Reagan appointees that were the most “extreme” Republican judges on the circuit courts).

22. See, e.g., David C. Vladeck, *Keeping Score: The Utility of Empirical Measurements in Judicial Selection*, 32 FLA. ST. U. L. REV. 1415, 1437–40 (2005) (suggesting that although judges

may capture little more than judicial disagreeability.²³ In that respect, there appears to be a sound norm that such disagreeability at the circuit court level is not a praiseworthy judicial attribute.²⁴ Even if the independence measure does capture non-ideological judging, it might be a poor test for Supreme Court appointees because as we will discuss in Part I.B.2, ideology plays an important role on the Court.

The citation test has more validity, but it is also controversial. The reliability of citation analysis as a measure of judicial quality has been much ventilated in earlier research. One problem associated with the measure is that citations may be negative, criticizing the cited opinion.²⁵ More citations may reflect longevity on the bench or some other feature rather than judicial quality. Citations vary in importance; some are centrally relied upon to reach decisions, whereas others are fodder for only a string cite. Citation decisions may be infected by judicial friendship (or animus), and they may be influenced by the choices of judicial clerks as much as the citing judges themselves.²⁶ Indeed, clerks may have largely written the cited opinions for which the judges are receiving credit in the tournament.²⁷

who vote together against the desires of their appointing president should be given credit for independence, the measure gives them no credit).

23. Professors Choi and Gulati conceded that this measure might really capture “cantankerousness and unwillingness to compromise as opposed to real independence.” Choi & Gulati, *Mr. Justice Posner?*, *supra* note 1, at 38. The willingness to author a separate opinion might also be a reflection of judicial self-centeredness. See Farber, *supra* note 14, at 1177.

24. See James J. Brudney, *Foreseeing Greatness? Measurable Performance Criteria and the Selection of Supreme Court Justices*, 32 FLA. ST. U. L. REV. 1015, 1031 (2005) (“[J]udges are often critical of any pronounced tendency to write separate opinions. They fear the erosion of institutional integrity that may result . . . [and] 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.”). Circuit court judges have written on the values of collegiality. See, e.g., Harry T. Edwards, *The Effects of Collegiality on Judicial Decision Making*, 151 U. PA. L. REV. 1639, 1640–41 (2003) (“[C]ollegiality mitigates judges’ ideological preferences and enables us to find common ground and reach better decisions.”); Deanell Reece Tacha, *Judges on Judging: The “C” Word: On Collegiality*, 56 OHIO ST. L.J. 585, 586 (1995) (“I nevertheless believe that there is a value in collegiality that affects the quality of judicial decisionmaking.”). Scholars consider collegiality a criteria that good judges should possess. See Gerhardt, *supra* note 2, at 358.

25. Professors Choi and Gulati take note of the possibility that some citations may be attributable to “outrageously” bad decisions and hence measure negative citations for a subset of their judges. Choi & Gulati, *Choosing the Next*, *supra* note 1, at 55. They found that negative citations did not explain the high citation rates of their top judges. *Id.* at 56–57.

26. William M. Landes, Lawrence Lessig & Michael E. Solimine, *Judicial Influence: A Citation Analysis of Federal Courts of Appeals Judges*, 27 J. LEGAL STUD. 271, 272–75 (1998) (setting out these and other potential criticisms of citation analysis in detail).

27. Professors Choi and Gulati have recognized the widespread belief that circuit court opinions are often written by clerks. They tried to isolate those judges most likely to draft their

Although all these criticisms are relevant concerns, none of them have been documented as seriously undermining the validity of the citation measure, which remains commonly used.

Underlying most of the common criticisms of citation analysis is the concern that an uncontrolled factor, independent of judicial quality, has infected the analysis. For example, a judge setting forth the first explication of standards on a procedural issue such as summary judgment may be repeatedly cited simply because the issue so commonly arises in litigation. Similarly, some types of cases are simply more common and therefore more likely to receive citations than other types of cases. Not all citations are “created equally”;²⁸ citation frequency may even be a matter of “just plain luck.”²⁹ This does not invalidate the quantitative analysis of citations, however. Such errors are likely to be randomly distributed throughout the judicial population and may therefore be considered statistical “noise.”³⁰ The presence of this feature tends to make it more difficult to find true statistical significance and consequently may add further confidence to results that find such significance.³¹ Nevertheless, the possibility of such random noise does caution against giving undue weight to any single study’s conclusion, especially if a sample size is relatively small.³²

Although citation analysis is plainly imperfect, like any measure, it offers very useful information about circuit court judges. It provides a more rigorous quantitative scale, in addition to the more common subjective evaluations that may speak more about the judgments of the evaluators than the quality of the judges themselves. Indeed, a “judge whose opinions are consistently useful to others is probably doing something right, whereas a judge whose opinions are rarely

own opinions. See Stephen J. Choi & G. Mitu Gulati, *Which Judges Write Their Opinions (and Should We Care)?* 3 (N.Y. Univ. Law Sch. Law & Econ., Working Paper No. 05-06, 2005), available at <http://ssrn.com/abstract=715062> (“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.”).

28. Vladeck, *supra* note 22, at 1432.

29. Farber, *supra* note 14, at 1178.

30. Richard A. Posner, *An Economic Analysis of the Use of Citations in the Law*, 2 AM. L. & ECON. REV. 381, 390 (2000).

31. For some discussion of this point, see *id.*

32. The Choi and Gulati results are limited to three years of opinions. This is not a criticism, as the data collection efforts required for such a study are daunting. However, it counsels some caution in interpretation, as a larger sample might yield different results, and judges’ quality may change over their tenure on the bench.

cited is probably performing badly.”³³ Indeed, circuit court judges have testified to the practical validity of this measure. In one survey, most of the judges interviewed “felt that the name on the opinion” cited by attorneys “did affect their decisionmaking at times.”³⁴ Although citation influence reflects one scale of judicial quality, akin to influence, as discussed below, the normative significance of this scale remains debatable.

Others have taken issue with Professors Choi and Gulati’s exclusive focus on circuit court judges in their rating system, at the expense of state court judges or politicians. These critics note that the highest-rated Supreme Court Justices have come from other sources.³⁵ This is an accurate point, but it does not demean Choi and Gulati’s effort. Today’s Court is composed solely of former circuit court judges, as the circuit courts have become the most common source for Supreme Court nominees.³⁶ The proposed tournament at least allows a screening for this category of appointees.³⁷

2. *Broader Usage Critiques.* Critics have also focused more generally on the use of the Choi and Gulati methodology. These criticisms questioned the wisdom of their chosen methodology notwithstanding the accuracy of its scales. In social scientific terms, the critiques questioned the measures’ validity rather than their reliability. In this view, the variables they measured do not address the key concerns regarding judicial quality. Even if the measures did capture the concept of judicial quality at the circuit court level, those metrics are not necessarily relevant to the qualities that are important at the Supreme Court level.

One of the most extensive critiques of Professors Choi and Gulati that blends methodology with usage comes from Larry Solum, who has produced his own “Solumonic” standards for judicial aptitude.³⁸ He argues that Choi and Gulati have avoided the difficult

33. Farber, *supra* note 14, at 1179.

34. KLEIN, *supra* note 5, at 94.

35. See Steven Goldberg, *Federal Judges and the Heisman Trophy*, 32 FLA. ST. U. L. REV. 1237, 1237 (2005); WERL, *supra* note 2, at 165–67.

36. See Lee Epstein, Jack Knight & Andrew D. Martin, *The Norm of Prior Judicial Experience and Its Consequences for Career Diversity on the U.S. Supreme Court*, 91 CAL. L. REV. 903, 905 (2003) (addressing the increased reliance on promotion from circuit courts).

37. Professors Choi and Gulati acknowledge this point and suggest the possibility of empirically measuring candidates from other fields as well. Choi & Gulati, *A Tournament*, *supra* note 1, at 319–20.

38. See Solum, *supra* note 2 (manuscript at 1–2).

question of identifying judicial virtues. In this regard, Professor Solum identifies the relatively uncontroversial “thin” judicial virtues as including incorruptibility, courage, temperament and impartiality, diligence and carefulness, intelligence, and craft.³⁹ He also sets forth thick virtues of justice and practical wisdom.⁴⁰ These standards roughly trace traditional subjective scales for judicial quality.⁴¹ Professor Solum suggests some approaches to screen for these virtues, including evaluation of opinions and the judges’ involvement in public life.⁴²

Professor Solum argues that the citation measure fails to capture these thin and thick virtues, suggesting that citations may be driven by convenience and thus that citation frequency is not necessarily a good guide to opinion quality.⁴³ He maintains not only that Professors Choi and Gulati measured the wrong factors but also that the very “selection of judges on the basis of measurable performance criteria would lead us away from true excellence.”⁴⁴ Still worse, he suggests that the proposal’s end result would be “awful,” with judges lacking the “virtues of integrity, wisdom, or justice.”⁴⁵ Given the results of the tournament, this criticism may be somewhat hyperbolic. Few would argue that Judge Posner is utterly lacking in these Solumonic values. Moreover, the citation measure seems reasonably associated with many of Solum’s standards, including the thin virtues of diligence, craft, intelligence, as well as (plausibly) the thick virtues of justice and practical wisdom. Unless one has a very cynical view of the process of citation, the measure probably captures some element of judicial virtue as identified by Professor Solum.⁴⁶

39. *Id.* (manuscript at 4–8).

40. *Id.* (manuscript at 8–10).

41. *See, e.g.*, HENRY J. ABRAHAM, JUSTICES, PRESIDENTS, AND SENATORS: A HISTORY OF THE U.S. SUPREME COURT APPOINTMENTS FROM WASHINGTON TO BUSH II, at 1–2 (5th ed. 2008) (suggesting eight standards including judicial temperament, knowledge of the law, and the ability to handle judicial power sensibly); ALBERT P. BLAUSTEIN & ROY M. MERSKY, THE FIRST ONE HUNDRED JUSTICES 50–51 (1978) (suggesting standards including analytical powers, knowledge, courage, and character).

42. Solum, *supra* note 2 (manuscript at 11).

43. *Id.* (manuscript at 15–17) (noting that other variables, such as the ease with which that authority can be located, influence the probability that a given authority will be cited).

44. *Id.* (manuscript at 1).

45. *Id.* (manuscript at 22).

46. It is possible that this is not the case, because of the nature of opinions receiving more citations, which we will discuss extensively below. *See infra* Part III.A.1.

With respect to his specific arguments, Professor Solum maintains that the quality of judges cannot be precisely quantified by any measure. Yet Professors Choi and Gulati do not profess that their system is perfect; as Professor Solum acknowledges, they only wish to “do better than the status quo.”⁴⁷ Although Solum claims that reliance on measurable criteria might worsen the system, he makes no effort to examine, much less defend, the status quo. nor does he offer any other than amorphous proposals to achieve the same objective as Choi and Gulati’s tournament,⁴⁸ or identify any particular judges who rate well or poorly according to his criteria.⁴⁹ Thus Solum offers insight into judicial virtue but does not delegitimize the tournament approach. Nevertheless, Professor Solum does reasonably argue that Choi and Gulati’s specific operationalization of judicial quality is arguably flawed. These arguments are considered in Part III.

Another criticism arises from the possibility that the rating system might be manipulated or “gamed” by circuit court judges maneuvering for a promotion.⁵⁰ For example, judges may readily raise their independence measure simply by writing concurring or dissenting opinions. The citation measure is not readily manipulable by the opinion-writing judge but certainly can be manipulated by friends or co-ideologues. Alternatively, presidents might manipulate the system by using defective quality ratings to shield otherwise ideological appointment to the Court.⁵¹ Professor Solum raises this criticism and discusses how each of the measures might be gamed.⁵² Yet while such strategic behavior could undermine the measure’s validity, the criticism lacks the necessary comparative element

47. Solum, *supra* note 2 (manuscript at 2).

48. *See id.* (manuscript at 11–12). His proposals range from the insignificant (Supreme Court appointees should be over thirty years old), *id.* (manuscript at 11), to the unhelpful (they should be screened for the “lawfulness” of their opinions), *id.* (manuscript at 12). Solum contends that “we have good reason to believe” that we can screen for his virtues, *id.*, but he offers no specifics on how to do so or how likely the government is to adopt any such specifics, should they be identifiable.

49. Absent such specification, the criteria for quality judges are not helpful. *See* Frank B. Cross, *Gay Politics and Precedents*, 103 MICH. L. REV. 1186, 1208 (2005) (book review) (observing that critics of the present system have produced only “amorphous” standards for judicial quality and have failed to offer examples of “quality judges,” such that their criticism offers little value to the confirmation process).

50. Marshall, *supra* note 2, at 122–23.

51. *Id.* at 131.

52. Solum, *supra* note 2 (manuscript at 18–20).

because the present system may also involve gaming one's promotion prospects. Strategic networking probably infects any selection system.

Under the status quo system, judges can game promotion by rendering decisions amenable to the potential appointing president. Professor Solum might find this type of behavior especially abhorrent because of its potential to compromise justice and the rule of law for selfish advantage. Empirical evidence does exist to suggest that such strategic behavior has occurred in district court decisions on the constitutionality of the Federal Sentencing Guidelines. The findings of an extensive study of Guideline rulings found that the promotion potential of district court judges was a statistically significant determinant of case outcomes.⁵³

Perhaps the most trenchant usage critique for the Choi and Gulati measures focuses on the special role played by the Supreme Court. The Supreme Court's unique role in the American political system undermines any attempt to extrapolate circuit court performance as a measure of the quality of a Supreme Court Justice.⁵⁴ Professors Choi and Gulati recognize that "[t]he best soldiers are not always the best leaders."⁵⁵ The shortcoming of their methodology for selecting leaders is most obvious with the productivity test, as discussed in Part I.B.1, but this shortcoming also affects their other standards to some degree as well.

The goal of judicial evaluation research may be to develop some objective test for Supreme Court appointments. Professors Choi and Gulati specifically comment that their effort was borne of "frustration with the current appointment process."⁵⁶ There is indeed a widespread belief that the Court's appointment system is somehow

53. Gregory C. Sisk, Michael Heise & Andrew P. Morriss, *Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning*, 73 N.Y.U. L. REV. 1377, 1488-93 (1998). Similar results were found by Mark A. Cohen. See Mark A. Cohen, *Explaining Judicial Behavior or What's "Unconstitutional" About the Sentencing Commission?*, 7 J.L. ECON. & ORG. 183, 198 (1991). Judges seeking promotion "have to toe the popular line or at least a line acceptable to the Senate Judiciary Committee." Pamela S. Karlan, *Two Concepts of Judicial Independence*, 72 S. CAL. L. REV. 535, 545 (1999) (footnote omitted).

54. See, e.g., Vladeck, *supra* note 22, at 1417 ("[T]he talents one needs to be a superior Supreme Court Justice are light years apart from those needed to stand out on the court of appeals, and thus even a stellar performance on the court of appeals is no predictor of success on the high court."); see also Goldberg, *supra* note 35, at 1242-43 (drawing the analogy between success in college football and failure at the professional level). James Brudney uses a comparison of Burger and Blackmun's records on the circuit court and Supreme Court to make this same point. See Brudney, *supra* note 24, at 1029-38.

55. Choi & Gulati, *A Tournament*, *supra* note 1, at 310.

56. Choi & Gulati, *Choosing the Next*, *supra* note 1, at 26.

broken and needs to be “fixed.”⁵⁷ This belief is not uncommon, as commentators have suggested that the confirmation process produces only mediocre judges⁵⁸ and is best described as a “mess.”⁵⁹ Others more modestly suggest that the confirmation process precludes the selection of “great” Justices and is biased in favor of “competent, noncontroversial jurists with a restrained understanding of the role of the federal judiciary.”⁶⁰ Prominent among the criticisms is that the selection process focuses unduly on judicial ideology.⁶¹

The ideological focus in appointments is not necessarily misplaced, however; it may be appropriate and inevitable,⁶² as argued by Terri Jennings Peretti.⁶³ She notes that presidents have generally appointed judges based on ideological compatibility and that senatorial confirmation has likewise been ideologically grounded.⁶⁴ This ideological influence is normatively appropriate, she believes,

57. See Cross, *supra* note 49, at 1206–09 (reviewing the criticisms of the current nomination process).

58. See, e.g., Bruce Fein, *A Court of Mediocrity*, A.B.A. J., Oct. 1991, at 74, 74 (stating that since Benjamin Cardozo, “subsequent Supreme Court nominees and appointees have been invariably pedestrian”).

59. See STEPHEN CARTER, *THE CONFIRMATION MESS: CLEANING UP THE FEDERAL APPOINTMENTS PROCESS* *passim* (1994); Stephen Carter, *The Confirmation Mess, Revisited*, 84 NW. U. L. REV. 962 *passim* (1990). Other descriptions use words like “mess, abysmal, broken, going in the wrong direction, and downright disorderly, contentious, and unpredictable.” 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, 1146 (2005) (footnotes omitted) (internal quotation marks omitted).

60. MARK SILVERSTEIN, *JUDICIOUS CHOICES: THE NEW POLITICS OF SUPREME COURT CONFIRMATIONS* 175 (updated college ed. 1994).

61. See Choi & Gulati, *Choosing the Next*, *supra* note 1, at 26 (“The genesis of this project lies in our frustration with the current appointment process. As best as we can tell, the entire focus in analyzing a candidate’s qualifications is on predicting their expected votes on a handful of issues.”).

62. Professors Choi and Gulati acknowledge this point to some degree, and argue that their measures are simply a tool to make politicians openly acknowledge ideological influence. *Id.* at 27–28. Given the nature of the process, though, this hardly seems necessary. It is an open secret that ideology largely determines Supreme Court appointments. Choi and Gulati go on to refer to it as an “informal norm,” *id.* at 37, and, in another work, they concede that “it is well recognized that politics in fact does play a large (and constitutional) role in both the nomination and confirmation of federal judges,” Choi & Gulati, *A Tournament*, *supra* note 1, at 301 n.5. It seems unlikely that either voters or senators seriously credit presidential claims that appointments are based on quality. See Brannon P. Denning, *Empirical Measures of Judicial Performance: Thoughts on Choi and Gulati’s Tournament of Judges*, 32 FLA. ST. U. L. REV. 1123, 1132 (2005) (“[P]erhaps [Choi and Gulati] simply overestimate the degree to which voters and Senators credit presidential claims of merit.”).

63. TERRI JENNINGS PERETTI, *IN DEFENSE OF A POLITICAL COURT* (1999).

64. *Id.* at 85–89.

because it yields a politically representative Court, albeit with some time lag.⁶⁵ Insofar as Justices reach decisions of political salience when the underlying law is indeterminate, such a connection is a defensible one. Thus, Professor Peretti contends that policy-motivated judges further the basic goals of democracy in a way that apolitical judges cannot.

The Supreme Court by its nature makes some discretionary, ideological judgments that are not resolved by technical legal acumen.⁶⁶ For example, giving content to the meaning of “due process” requires discretionary judgment. Although Solumonic “practical wisdom” is relevant to this discretionary judgment, there are nonetheless ideological elements central to the meaning. Ample empirical research demonstrates that ideological decisionmaking is commonplace at the Court.⁶⁷ Given this finding, it appears that “[l]egal quality matters least on supreme courts.”⁶⁸ To the extent that ideological judgments are an important part of a Justices’ role, it is more sensible that political ideology influences their selection. Of course, an utterly ideological process, without any regard for judicial aptitude of the Solumonic sort, would be undesirable. But there is no evidence that such an unfortunate situation prevails.⁶⁹

Lee Epstein, Jeff Segal, Nancy Staudt, and Rene Lindstadt conducted a study in which they considered the perceived qualifications of the appointees, as well as their ideologies.⁷⁰ They find that relative qualifications are a statistically significant factor in

65. *Id.* at 100–01.

66. See CHARLES L. BLACK, JR., *THE PEOPLE AND THE COURT: JUDICIAL REVIEW IN A DEMOCRACY* 174 (1977) (observing that such legal acumen is “not a sufficient condition . . . for dealing competently with questions of constitutional law” when political judgment is inevitable); Brudney, *supra* note 24, at 1044 (contending that the “real world seeks considerable value” in the interplay of law and politics in the Supreme Court).

67. For a review of this evidence, see PERETTI, *supra* note 63, at 102–11. The classic work of empirical research on this question is JEFFREY A. SEGAL & HAROLD J. SPAETH, *THE SUPREME COURT AND THE ATTITUDINAL MODEL* (1993). Their findings have been confirmed by considerable subsequent research. A meta-analysis of the empirical research shows consistent ideological associations in decisionmaking, which are particularly pronounced at the Supreme Court level. Daniel R. Pinello, *Linking Party to Judicial Ideology in American Courts: A Meta-analysis*, 20 *JUST. SYS. J.* 219, 243 (1999).

68. Cross, *supra* note 49, at 1211. This claim does not depend on the Justices being particularly willful and law-disregarding; it is a feature of the Court taking the sorts of close cases for which traditional legal materials cannot answer the questions presented. See *id.*

69. See *id.* (noting that the evidence shows that judges compromise their ideological preferences when the law is plainly to the contrary).

70. Epstein et al., *supra* note 59, at 1160 fig.3, 1162 fig.4.

Senate votes for nominees, both by themselves and when linked with ideological concerns.⁷¹ Moreover, they find that a very highly qualified nominee is virtually assured of confirmation, regardless of ideological concerns.⁷² This aspect of the process might be improved, though, by the proposed tournament if it effectively captured qualifications more precisely than contemporary conventional subjective measures.

Contrary to claims sometimes made, it is unclear that the assertedly growing importance of judicial ideology in the appointment process has reduced the quality of Supreme Court Justices in recent years. Michael Comiskey has analyzed the estimated quality of Supreme Court Justices over the years, through a broad survey of professors of law and political science.⁷³ He found that the average quality of selections since 1967 was actually higher than that of earlier selections.⁷⁴ Although his measures were purely subjective, there is no empirical evidence whatsoever demonstrating that recent appointments were of lower quality.

It may be that there is no methodology that can precisely predict the future quality of a Supreme Court Justice.⁷⁵ This does not undermine the effort to identify such a disciplined methodology of quality, though. One cannot know until one tries, and the effort may provide some useful information on judicial quality. Moreover, the rating of circuit judges has a variety of benefits, unrelated to Supreme Court appointment, which are addressed below.

71. *Id.* at 1171. The results of the study show that ideological agreement is significant, that qualifications are significant, and that qualifications are even more significant when the appointee is ideologically more distant from the voting senator. *Id.* at 1171–72.

72. *See id.* at 1170; *see also* KEVIN T. MCGUIRE, UNDERSTANDING THE U.S. SUPREME COURT 38–40 (2002) (reviewing recent nominations and finding that great importance was placed on a nominee’s qualifications).

73. MICHAEL COMISKEY, SEEKING JUSTICES: THE JUDGING OF SUPREME COURT NOMINEES 18 (2004).

74. *Id.* at 86–87. Comiskey’s results provided some support for the suggestion that modern confirmation prevents the appointment of “great” Justices because none of the modern Justices were categorized as great. *See id.* at 98–99 (observing that Antonin Scalia, the highest-rated appointee since 1967, received a rating significantly below the “greatness” cutoff). This may simply reflect a reluctance of the evaluators to ascribe greatness to their contemporaries, whose decisions have not yet stood the test of time. *See Cross, supra* note 49, at 1210 (“[P]erceptions of greatness appear only after time.”). Moreover, this effect is more than counterbalanced by the apparent ability of the current process to screen out unusually bad Justices, evidenced by the higher mean evaluation of quality.

75. Michael J. Gerhardt, *The Art of Judicial Biography*, 80 CORNELL L. REV. 1595, 1641–45 (1995) (book review) (arguing that it is virtually impossible to predict the greatness of a Justice in advance).

The research on the prevailing selection and confirmation methods does not demonstrate the optimality of the selection process, and one might argue that quality measures have at least some value in choosing the best-quality judges within a given ideological group. This point has some validity, but it means that a more refined test for quality is necessary. Such a test requires an understanding of the different types of judging and the meaning of judicial quality, as well as the ability to capture that meaning. This is a daunting task. Before turning to this analysis, we review other published research on the evaluation of judges.

C. *Other Research Assessing Judges*

Several other empirical studies have sought to measure the quality of judges empirically, and they are reviewed in this Section. Like Professors Choi and Gulati's study, most of these studies have focused on the circuit court judiciary, and they have commonly used some measure of citation frequency as their test for quality. Research in this field generally traces to a 1976 article by Professors William Landes and Richard Posner, setting out an economic model of precedent.⁷⁶ Landes and Posner theorized that the importance of a precedent is represented by the number of times that it is cited. The main finding of the article was that the influence of precedents depreciated over time, as expected. Posner has built on this article to further analyze the value of citation analysis in various contexts.⁷⁷

Professors Landes, Lawrence Lessig, and Michael Solimine conducted an elaborate citation analysis of circuit court judges to measure "judicial influence."⁷⁸ They measured outside circuit citations for judges sitting on the circuit courts from 1992 to 1995. On their scale of influence, Judge Posner was first and Judge Easterbrook third, with Judge Selya of the First Circuit coming in second.⁷⁹ Judge Selya jumped to the top of the list for total citations.⁸⁰ Using different data, their results essentially conform to the findings of Choi and Gulati.

76. See William M. Landes & Richard A. Posner, *Legal Precedent: A Theoretical and Empirical Analysis*, 19 J.L. & ECON. 249, 271 (1976).

77. Posner, *supra* note 30 *passim*.

78. Landes et al., *supra* note 26, at 276–79.

79. *Id.* at 288.

80. *Id.* at 298.

Professors David Klein and Darby Morrisroe conducted a major study on the prestige and influence of circuit court judges, using published opinions between 1987 and 1990,⁸¹ roughly parallel to the years covered by Professors Choi and Gulati. Their basic standard was the explicit invocation of particular judges by name in others' opinions.⁸² This represents a possible—but not definitive—direct measure of the respect accorded to judges by their fellow judges.⁸³ Most of the circuit court judges studied were not explicitly invoked by name, but others were named more frequently. The results of this study generally correspond to those of Choi and Gulati's study: Judges Posner and Easterbrook were third and fourth in their prestige score (the first-ranked judge might be ascribed to his Sentencing Commission service rather than opinions, whereas the judge ranked second, Stephen Breyer, was appointed to the Supreme Court in the interim).⁸⁴ Klein and Morrisroe also conducted a sensitivity test and found that their results were significantly associated with other measures such as citations and quality evaluations by practicing attorneys, but not with ABA ratings.⁸⁵

These prestige rankings received some validity confirmation in subsequent empirical research. Professor Klein went on to identify cases announcing new legal rules in the areas of antitrust, search and seizure, and environmental law that were decided during the 1980s and 1990s.⁸⁶ He then tested to see if subsequent circuit decisions adopted that rule and sought to determine the factors that predicted such an adoption. His judicial prestige measure, along with other factors, was a statistically significant determinant of subsequent adoptions.⁸⁷ Thus, invocations appear to be a meaningful measure of prestige that influences the subsequent course of the law.

A more recent analysis by Professors Jeffrey Berger and Tracey George has compared circuit court judges on a test of “judicial

81. David Klein & Darby Morrisroe, *The Prestige and Influence of Individual Judges on the U.S. Courts of Appeals*, 28 J. LEGAL STUD. 371, 379 (1999).

82. *Id.* at 376. This invocation measure was used because such a citation is unnecessary to the citing opinion, except insofar as the naming of a judge lends some persuasive import to the cited decision. *Id.* at 375–76.

83. *Id.* at 376.

84. *Id.* at 381.

85. *Id.* at 384–85.

86. KLEIN, *supra* note 5, at 40–41.

87. *Id.* at 78–79. Other significant determinants included the relative ideology of the judges and whether the decision was from the same circuit, confirming both the political and the legal model of judicial decisionmaking. *Id.* at 79.

entrepreneurship.”⁸⁸ This study examined Supreme Court citations to circuit court decisions in the 1994 to 2002 terms.⁸⁹ As in much of the quantitative research, Judges Easterbrook and Posner were the most-cited circuit court judges.⁹⁰ In general, the research on citations produces consistent results even over time and with different measures, at least at the top of the list of most-cited judges. This adds some credibility to the conclusions drawn about the “best” circuit court judges by this measure. The conclusions of the research appear quite robust. Choi and Gulati have observed that Judge Posner “is cited more by his colleagues on the Seventh Circuit, by other circuit judges, by law professors, and by the Supreme Court” than any other judge, and typically by a substantial amount.⁹¹

The existing research is largely internally consistent and supportive of Professors Choi and Gulati’s conclusions, which may be unsurprising given the general similarity of measuring tools it has used. Having reviewed this research on circuit court judges, we turn to our own elaboration of the analysis. In Part II, we add an important set of new data that measures the fate of the judges’ opinions in the Supreme Court. Once this is done, we can reevaluate the existing measures of judicial quality.

II. EVIDENCE FROM SUPREME COURT REVIEW

Commentators have occasionally suggested that circuit court judges might be evaluated based on the frequency with which the Supreme Court reversed their decisions. To date, however, this measure has not actually been implemented. This Part provides such data to implement this test of circuit court judges based on Supreme Court review. Although the success of judges’ opinions before the Supreme Court should not be considered the exclusive measure of circuit court judge quality, it is a legitimate factor to be considered.

88. Jeffrey A. Berger & Tracey E. George, *Judicial Entrepreneurs on the U.S. Courts of Appeals: A Citation Analysis of Judicial Influence* 2–3 (Vanderbilt U. Law Sch. Law & Econ., Working Paper No. 05-24, 2005), available at <http://ssrn.com/abstract=789544>.

89. *Id.* (manuscript at 3).

90. *Id.* (manuscript at 15).

91. Choi & Gulati, *Mr. Justice Posner?*, *supra* note 1, at 28–29. The Posner effect may be biased by his extensive outside writings as a public intellectual.

A. *Significance of Supreme Court Review*

The relative success of circuit court judges' decisions on Supreme Court review might seem like an obvious standard for their evaluation. On the other hand, a critic of this standard might suggest that the decision reversing the circuit court judge could also be erroneous and that the best circuit court judges are more likely to "get it right" than the Supreme Court on review of their rulings.⁹² Yet this criticism inevitably undermines every attempt to measure judicial quality and thus is not dispositive. There is no objective external test for the "legal correctness" of a decision or the "legal quality" of an opinion available to researchers.⁹³ Thus, while imperfect, reversal by the Supreme Court remains one reasonable approach to evaluating judicial quality.

Indeed, the Choi and Gulati measures and other citation tests analogously rely on a presumption that later citing judges reliably reflect quality assessment. A similar presumption for the high Court is equally fair and perhaps preferable because the Court's decisions *are* the law of the land. The Supreme Court hears many fewer cases and consequently has the advantage of having considerably more time to evaluate the legal issues. The Court often has far better legal and other information on which to ground its decisions.⁹⁴ Moreover, circuit courts may be regarded as agents of the Supreme Court,⁹⁵ so it seems appropriate to consider the evaluations of their principal.⁹⁶ Finally, Professors Landes, Lessig, and Solimine claim that the

92. See Vladeck, *supra* note 22, at 1435 (expressing doubt that "the Supreme Court always has the better argument").

93. See Cross, *supra* note 20, at 1499–500 (discussing the "difficulty in independently verifying the correctness of a decision").

94. Supreme Court cases receive far more information from *amici* than those at the circuit court level. Similarly, the litigants themselves typically devote more time and resources to briefing cases at the Supreme Court level. The Court also has the benefit of more support from clerks and the opportunity to choose the very best clerks from the circuit court level, after they have proved their adeptness in a prior circuit court clerkship.

95. See, e.g., Donald R. Songer, Jeffrey A. Segal & Charles M. Cameron, *The Hierarchy of Justice: Testing a Principal-Agent Model of Supreme Court–Circuit Court Interactions*, 38 AM. J. POL. SCI. 673, 673 (1994) (addressing the principal-agent relationship between the courts).

96. Brannon Denning seemingly disputes this and argues that "[a]n 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." Denning, *supra* note 62, at 1141. This is a potential shortcoming of the measure, but he offers no argument for how frequently this occurs, and its occurrence would again fall within the category of statistical noise, discussed above, which affects all conceivable quantitative metrics of evaluation.

Supreme Court “rarely takes cases” from top-ranked judges because those judges tend to “get things ‘right,’”⁹⁷ though this claim is still an open question.

Another more trenchant critique of the Supreme Court review standard lies in the Court’s ideological decisionmaking. It is beyond dispute that the Justices are influenced by their political ideologies in ruling on the cases that they choose to hear.⁹⁸ Consequently, the Court’s frequent reversal of a particular judge’s decisions might reflect nothing more than ideological incompatibility between the Court’s majority and the judge or opinion under review.⁹⁹ If so, reversals would reflect little about the actual legal quality of the circuit court judge’s opinions. Indeed, the association may be contrary to the correct measure of legal quality to the extent it is driven by ideology rather than legal norms or principles. One author suggests that the reversal rate for the Ninth Circuit was high because that court was “too law-abiding” under existing precedents at a time when the Rehnquist Court was devoted to changing the existing law.¹⁰⁰ Yet rigorous empirical analyses have demonstrated that the Court is influenced by legal norms as well as ideological considerations.¹⁰¹ The circuit courts annually render hundreds or even thousands of opinions that are ideologically contrary to the justices’ preferences. Most of these decisions are not reversed, suggesting that there remains some additional significance to a reversal beyond mere ideological incompatibility. Further, we can empirically examine the degree to which the ideological bias taints our results.

One can imagine other criticisms of the Supreme Court review standard. Circuit judges are individuals responsible for hundreds of

97. Landes et al., *supra* note 26, at 325–26.

98. *See supra* note 67.

99. *See* Choi & Gulati, *A Tournament*, *supra* note 1, at 307 (noting that this measure “may unfairly penalize judges with different political views from those on the Court”); Posner, *supra* note 4, at 1273 (“[M]any reversals by the Supreme Court reflect ideological differences . . .”).

100. Michele Landis Dauber, *The 9th Circuit Follows*, LEGAL TIMES, Aug. 19, 2002, at 36. This conclusion might be challenged, though, given the persistence of the circuit’s high reversal rate even after the Rehnquist Court had worked to change the law. In addition, the Ninth Circuit had an unusually high rate of unanimous reversals, joined by liberal Justices, which suggests that this is not the true explanation. *See* Marybeth Herald, *Reversed, Vacated, and Split: The Supreme Court, the Ninth Circuit, and the Congress*, 77 OR. L. REV. 405, 435–36 (1998).

101. *See* Stefanie A. Lindquist & David E. Klein, *The Influence of Jurisprudential Considerations on Supreme Court Decisionmaking: A Study of Conflict Cases*, 40 LAW & SOC’Y REV. 135, 148 (2006) (finding that, independent of ideology, the Court is more likely to follow the reasoning process adopted by the majority of circuits and more likely to adopt the position of more prestigious circuit court judges).

opinions per year, and the Court reviews a small percentage of these rulings. Many incorrect circuit court rulings may go unexamined by the Supreme Court, which is not a court of “error correction.”¹⁰² Reversal may therefore not capture the presence of a legal error below. A Supreme Court reversal of a given case may simply be due to “serendipity.”¹⁰³ One answer to this critique is that the Supreme Court selectively reviews the most important decisions rendered by circuit courts, making it reasonable to use these important decisions as a metric.¹⁰⁴ Moreover, to the degree that evaluation of circuit court judges is employed as a standard for Supreme Court appointments, it seems appropriate to consider the fate of their decisions at the Supreme Court level.

Another possible bias of reversal measures relates to Professors Choi and Gulati’s productivity measure for the number of opinions written. Circuit court judges who write more opinions have a correspondingly greater likelihood of Court reversal, but they also have proportionally more opportunities for Court affirmances. Including credit for affirmances helps to moderate this criticism of a reversal measure. We do not suggest that the Supreme Court review measure is the sole test for circuit court judicial quality or even a “gold standard” measure. We contend only that it is an important metric that has not been previously considered and that may yield insight into the quality of circuit court judges. In this regard, we note that Judge Posner has suggested that a judge’s “reversal rate” serves as a proxy for performance.¹⁰⁵ Moreover, as Sections B and C will show, the Supreme Court review standard substantially informs the meaning of the existing measures of circuit court judicial quality, including the Choi and Gulati measures. Thus it is valuable as a supplementary measure.

102. E.g., Michael Abramowicz & Maxwell Stearns, *Defining Dicta*, 57 STAN. L. REV. 953, 1047 n.257 (2005) (observing that “error correction is not a sufficient basis” for the Court to take a case); Kermit Roosevelt III, Essay, *Light from Dead Stars: The Procedural Adequate and Independent State Ground Reconsidered*, 103 COLUM. L. REV. 1888, 1916 (2003) (“[T]he Supreme Court does not ordinarily engage in error-correction.”).

103. Vladeck, *supra* note 22, at 1435. However, this is another factor that amounts only to statistical noise and is refuted somewhat by the nature of the results in Part II.C, which suggest a systematic effect.

104. This conclusion might be contested. Because the Court takes the most difficult cases, it may not screen for the most blatant legal errors in easier cases. The latter may be left to correction by lower courts.

105. Posner, *supra* note 4, at 1259.

B. *Data and Methods*

We gathered data on Supreme Court reversals from Supreme Court opinions published between 1989 to 2000. Although these data are not perfectly contemporaneous with data from the Choi and Gulati study, the Klein and Morrisroe study, and other sources, this limitation is by no means fatal—the underlying premise of any search for the “best” judge is that judicial quality of individuals is relatively invariant across time. If individuals’ judicial quality varies significantly over time, it renders it impossible to presume that the best contemporary judges will also be the best judges in future years. Some evidence does exist to suggest that judicial quality is consistent over time, given the relatively consistent ratings of Judges Posner and Easterbrook in the research described in Part I.C, which covered multiple years.

For each of the Supreme Court opinions, we coded the identity of the circuit court judge whose opinion was being reviewed by the Court, excluding *en banc* decisions. This identified the author of the opinion being reviewed, the judges who signed on to the opinion, and any judges who dissented from the opinion. The judges were broken down into six categories:

1. Authored opinion affirmed by the Supreme Court
2. Joined opinion affirmed by the Supreme Court
3. Dissented from opinion affirmed by the Supreme Court
4. Authored opinion reversed by the Supreme Court
5. Joined opinion reversed by the Supreme Court
6. Dissented from opinion reversed by the Supreme Court

This approach enables more elaborate testing of the fate of circuit court decisions than an exclusive reliance on reversals. From these different scales, we can assess the success of individual circuit court judges on Supreme Court review.

C. *Results*

This Section sets out the quantitative results of our analysis of the opinions of circuit court judges on Supreme Court review. We provide data for individual judges who were most often reversed and most often affirmed and who most often dissented from decisions reviewed by the Court. We also provide a composite measure to capture the totality of these effects. Finally, we also examine the association between these Supreme Court review metrics and

preexisting circuit judge ratings, such as the Choi and Gulati quality scale.

We began the analysis by identifying the circuit court judges who were most frequently reversed by the Supreme Court during the time period covered in our data. The mean frequency for judge reversals was 2.04. Table 1 reports the five judges most frequently reversed.

Table 1. Most Frequently Reversed

Reinhardt (9th)	14
Pregerson (9th)	9
Fletcher (9th)	8
Chapman (4th)	7
J. Gibson (8th)	7

Unsurprisingly, judges on the Ninth Circuit predominate the list. It is widely known that the Supreme Court has reversed many decisions from this circuit.¹⁰⁶ The reason for this is unclear; it may reflect an ideological effect, a possibility that is examined below. None of the very best Choi and Gulati judges were among those most frequently reversed by the Court.

Next we considered the frequency of reversal for the highest scoring judges on the Choi and Gulati quality scale. Table 2 reports the number of Supreme Court reversals for the top scorers on this quality scale.¹⁰⁷ Judge Lynch is not included in this list because she joined the circuit court midway through the Supreme Court review period that we considered.

Table 2. Choi and Gulati Winners and Reversal

Posner (7th)	4
Easterbrook (7th)	5
Selya (1st)	5

The Choi and Gulati tournament winners all suffered a significant number of reversals. Although they were not among the most reversed judges, they did each experience a reversal frequency

106. See, e.g., Jerome Farris, *Judges on Judging: The Ninth Circuit—Most Maligned Circuit in the Country—Fact or Fiction?*, 58 OHIO ST. L.J. 1465, 1465 (1997) (noting the high reversal rate for Ninth Circuit opinions reviewed by the Supreme Court).

107. These are among the top five judges as measured by Professor Choi and Gulati's quality measure of number of outside citations. See Choi & Gulati, *Mr. Justice Posner?*, *supra* note 1, at 28.

about twice that of the median circuit court judge. Of course, this result could stem from these judges' propensity for legal entrepreneurship or innovation.

Judicial quality might also be measured by the number of a judge's decisions affirmed by the Supreme Court. Because these affirmances generally involved particularly salient or controversial cases in which the circuits were split, they could provide a good measure of a judge's ability to reach the "right" decision in the sorts of cases that are heard by the Supreme Court. The mean affirmance frequency was 1.26 for the period analyzed. Table 3 reports the frequency of affirmances for the most-affirmed circuit court judges.

Table 3. Most Frequently Affirmed

Posner (7th)	8
Higginbotham (5th)	8
Sloviter (3d)	6
Widener (4th)	6
Winter (2d)	5
Seymour (10th)	5
Bowman (8th)	5

The next table takes the Choi and Gulati top scorers and considers the number of their opinions that were affirmed by the Supreme Court.

Table 4. Choi and Gulati Winners and Affirmance

Posner (7th)	8
Easterbrook (7th)	4
Selya (1st)	0

Judge Posner is at the top of the list of affirmances, which supports the suggestion that he is the top circuit court judge. Judge Easterbrook is also relatively high on the list, well above the median judge, but Judge Selya had no opinions affirmed during our time period. These results for Supreme Court affirmances are somewhat consistent with the Choi and Gulati findings.

Although these affirmance lists provide valuable information, they also have limitations. A circuit judge can only be affirmed when authoring a majority opinion. Consequently judges who are ideologically out of step with most of their circuits might be undercounted on affirmances, because they would be unable to

command majorities. For example, a conservative judge on a predominantly liberal circuit might find it difficult to write opinions affirmed by the Supreme Court, because such a judge would find it more difficult to command a majority on the panel. This effect may be operating for a conservative judge on the Ninth Circuit, for example. This flaw is corrected somewhat by the analysis of dissenting opinions we undertake immediately below.

Perhaps the best single measure of judicial quality is found in dissents. A judge who dissented from a circuit court opinion reversed by the Supreme Court “got it right,” even when the panel majority “got it wrong.” Unfortunately, this measure is compromised by the relative infrequency of dissents from circuit court panels,¹⁰⁸ and thus the sample of these cases is smaller.

Table 5. Most Frequent Dissenters from Reversed Opinions

Kozinski (9th)	6
Wallace (9th)	6
Rymer (9th)	4
Kleinfeld (9th)	4

As hypothesized, this scale is dominated by Ninth Circuit judges, who had a particularly large number of reversed cases from which they had the opportunity to dissent. The scale is led by the more conservative judges on the circuit. From the Choi and Gulati list, Judge Posner dissented from one reversed opinion, and Judge Easterbrook dissented from two reversed opinions. Both Posner and Easterbrook also dissented from one opinion that was subsequently affirmed by the Supreme Court.

To obtain an overall measure of success before the Supreme Court requires the combination of these results. We created a composite variable to reflect the cumulative success of judges before the Supreme Court, labeled *Reviewscore*. The weighting of this variable is somewhat arbitrary, but we gave greater weight to opinion authors and dissenters (both positive and negative) than to those who simply joined an opinion that was reviewed.¹⁰⁹ Because the Supreme

108. For a review of the practice of dissenting on circuit courts, see generally VIRGINIA A. HETTINGER, STEFANIE A. LINDQUIST & WENDY L. MARTINEK, *JUDGING ON A COLLEGIAL COURT* (2006). The overall probability of a dissenting opinion on the circuit courts is about 4 percent. *Id.* at 66 tbl.2.

109. *Reviewscore* is based on the following weights:

1. Authored opinion affirmed by the Supreme Court = 2

Court tends to reverse more decisions than it affirms, the average score was a negative one. The mean *Reviewscore* for the judges was -2.18. Table 6 reports the top five scores of circuit court judges on this metric.

Table 6. Top Judges on Reviewscore

Mansmann (3d)	9
Rosenn (3d)	8
Kozinski (9th)	8
Wallace (9th)	7
Buckley (D.C.)	7

These results provide a cumulative measure of the success of circuit judge decisions before the Supreme Court, considering authored opinions, joined opinions, and dissents as compared with results at the Court. Although our weighting process might be modified, changes are unlikely to produce dramatic results in the relative standing of the judges. Table 7 reports the results of the top judges on the Choi and Gulati measure with our more comprehensive *Reviewscore* variable.

Table 7. Choi and Gulati Winners and Reviewscore

Posner (7th)	6
Easterbrook (7th)	1
Selya (1st)	-10

With his higher number of affirmances, Judge Posner does much better than the median judge and is near the circuit court judges with the top *Reviewscores*, tied for sixth. Judge Easterbrook is only slightly better than the median judge, and nowhere near the top judges on this scale. On the Supreme Court review measure, Judge Selya does poorly, with a score that is well below the median.

Our Supreme Court review measure is not wildly contrary to Professors Choi and Gulati's findings, but neither does it confirm them. Their top judge, Posner, does well on the Supreme Court

2. Joined opinion affirmed by the Supreme Court = 1
3. Dissented from opinion affirmed by the Supreme Court = -2
4. Authored opinion reversed by the Supreme Court = -2
5. Joined opinion reversed by the Supreme Court = -1
6. Dissented from opinion reversed by the Supreme Court = 2

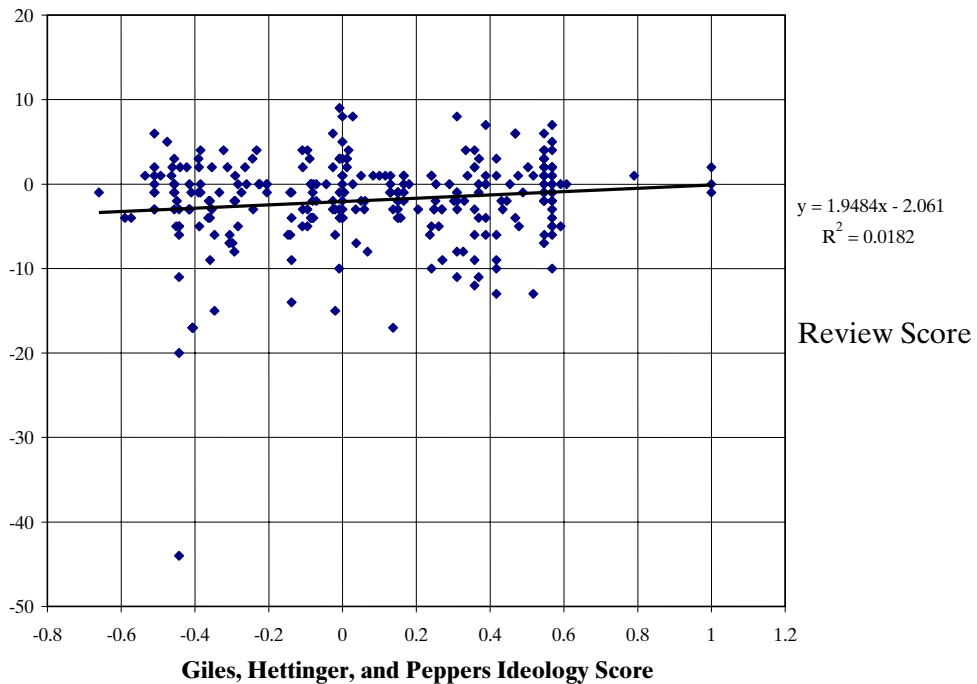
This system gave particular credit to judges whose opinions were affirmed by the Supreme Court or those who dissented from circuit court opinions later reversed by the Supreme Court.

review measure, but he is not at the very top of the measure. Their clear second judge, Easterbrook, does only modestly well on the Supreme Court measure. Another top-scoring judge in the preexisting research, Selya, actually does quite poorly on our measure.

Because the use of Supreme Court review has been criticized for potential ideological biasing of quality, we tested the data for the possibility of ideological skewing of the review score ratings. To do so, we first assigned a point estimate of ideology for each circuit court judge, computed by Professors Michael Giles, Virginia Hettinger, and Todd Peppers.¹¹⁰ On this scale, higher scores are associated with higher measures of a judge's conservatism. These scores were tested as a determinant of our cumulative *Reviewscore* variable. The results are graphically depicted in the following figure.

110. See Micheal W. Giles, Virginia A. Hettinger & Todd Peppers, *Picking Federal Judges: A Note on Policy and Partisan Selection Agendas*, 54 POL. RES. Q. 623, 624 (2001). The estimates for individual judges were based on presidential ideology plus consideration of senatorial ideology when senatorial courtesy was relevant to the appointment. *Id.* at 626–27.

Figure 1. Ideology and Reviewscore



The regression line is slightly ascending, which indicates that conservative circuit judges tend to have slightly higher *Reviewscores* than liberal judges, but the association does not approach statistical significance. Moreover, the R^2 term was less than .02, which demonstrates that any association was a very slight one that would not materially bias the use of this measure in evaluation. There was some “Ninth Circuit effect” in our results. But there is no obvious reason why the Supreme Court would have a vendetta against this circuit, so the results presumably display the Court’s evaluation of the correctness of that circuit’s decisions. By contrast, Professors Choi and Gulati’s citation analysis apparently does contain a substantial ideological bias.¹¹¹ Hence, this potential bias counsels in favor of

111. See WERL, *supra* note 2, at 174 (reporting a statistically significant association between appointment by a Republican president and higher citation counts).

considering Supreme Court review in the evaluation of circuit court judges.

Although we have compared Professor Choi and Gulati's top scorers with the results of Supreme Court review, a more rigorous statistical analysis is necessary to assess the association of their scores with Supreme Court results. Consequently, we now undertake regressions comparing the scores from Choi and Gulati's citation count measure, which they described as their measure of judicial quality. Table 8 reports the results of simple regressions using this measure of quality as an independent variable, and using *Reviewscore*, reversals, and affirmances as the dependent variables. The t-score for each association is in parentheses after its coefficient.

Table 8. *Choi and Gulati Quality and Supreme Court Success*

	<i>Reviewscore</i>	Reversed	Affirmed
C&G Quality	-.0003	.001	.002
R ²	0.00	0.02	0.11
N	58	58	58

Higher quality on Choi and Gulati's outside circuit citations actually corresponds with a higher rate of reversal and a lower *Reviewscore*, though neither association is statistically significant. A higher quality score also translates into a higher rate of affirmance by the Supreme Court, and only this association is statistically significant. The quality judges on the Choi and Gulati measure appear to be fairly aggressive in their decisionmaking, provoking more frequent Supreme Court review. They are relatively successful in achieving higher numbers of affirmances, but they also suffer more losses than the average judge.

To explore the association between measures of judicial quality, we expand our analysis to consider other research. We can produce the correlations between various measures, and we employ the *Reviewscore* from our data; the quality score from the Choi and Gulati study; the prestige score from the Klein and Morrisroe study;¹¹² the ratings of judges summarized in the Almanac of the Judiciary, also from the Klein and Morrisroe study; and the judges' Supreme Court citation rate (SCcite), from the Berger and George study.¹¹³ Table 9 presents an intercorrelation matrix showing the degree to which each of these measures correlates with each other measure.

112. Klein & Morrisroe, *supra* note 81, at 375–76.

113. Berger & George, *supra* note 88 (manuscript at 3).

Table 9. Intercorrelation of Judge Scores

	<i>Reviewscore</i>	Quality	Prestige	Almanac	SCcite
<i>Reviewscore</i>	-				
Quality	.0907	-			
Prestige	.3136	.8950	-		
Almanac	.5980	.2168	.3460	-	
SCcite	.1090	.8874	.8351	.2454	-

The clearest finding is that quality, prestige, and SCcite are highly correlated, which strongly suggests that they are capturing some underlying measure of judging. The relative lack of association with *Reviewscore* and the Almanac rating, though, suggests that these measures may capture only one dimension of judicial quality. Interestingly, the ratings of judges by lawyers in the Almanac correlate more closely with *Reviewscore* than with any of the other variables. Perhaps lawyers and the Supreme Court are seeing judicial qualities that are not measured by the circuit court citation and invocation scales.

We are hesitant to ascribe too much significance to our Supreme Court review results. We do not claim that they are a pure measure of judicial quality or that they are ambiguously superior to other metrics. Judge Reinhardt of the Ninth Circuit had the worst *Reviewscore* of any judge in our data, at -44. Whatever one thinks of his opinions, we suspect that few would regard him as clearly the *worst* circuit court judge sitting on the bench. His outside circuit citations in the Choi and Gulati study exceeded the median. Instead, we offer these Supreme Court-based scores as one additional dimension of quality to be considered in the evaluation. The relative scores of the judges are surely explained in part by the different approaches to judging discussed in Part III.

III. JUDICIAL TYPES COMPARED AND CONTRASTED

The use of a single quality scale or even multiple scales combined together implicitly presumes that there is one measure of judicial quality that need only be identified and quantified. It is possible, however, that there are multiple types of judges with multiple types of virtues and shortcomings, with no one type clearly preferable to others. In this Part, we consider the possible typologies of circuit court judges and how they may be seen in the quantitative empirical measures used to evaluate the judges.

Professor Solum discussed the many features that factor into assessments of judicial quality, but these features can be very difficult to capture empirically. Professors Choi and Gulati concede that their empirical results miss important considerations, “such as a judge’s propensity to be fair, do justice, exercise judgment, and demonstrate judicial temperament.”¹¹⁴ The Supreme Court review measures get at these considerations only very indirectly and imprecisely. The various measures, though, can help us identify distinct types of judges. These identifications can then be used, when combined with normative arguments, to evaluate different judges for different purposes.

A. *Judicial Types*

Circuit court judges may have different theories of judging, and these theories can influence our study of quality. The classic study of courts of appeals was conducted by Professor J. Woodford Howard.¹¹⁵ Based on surveys of Second, Fifth, and D.C. Circuit judges, he identified two basic judicial types. He characterized a minority of judges as “innovators” who “felt obliged to make law whenever the opportunity occurs.”¹¹⁶ One judge declared that the best part of being a judge was “launching new ideas.”¹¹⁷ By contrast, a slightly larger minority of judges, whom Howard called “interpreters,” believed that “judicial lawmaking should be held to a minimum.”¹¹⁸ A judge of this category declared that he “should leave innovation within the confines of the particular case and leave wholesale innovation to the legislature.”¹¹⁹ The majority of judges, called “realists,” took a hybrid position less innovative than that of the innovators but more innovative than that of the interpreters.¹²⁰

Professor David Klein’s subsequent survey of circuit court judges also found different values in the judiciary. Most judges regarded “legally correct decisions” as very important but put widely varying importance on having a “coherent, uniform law” or producing

114. Choi & Gulati, *Mr. Justice Posner?*, *supra* note 1, at 42.

115. See J. WOODFORD HOWARD, JR., *COURTS OF APPEALS IN THE FEDERAL JUDICIAL SYSTEM* (1981).

116. *Id.* at 160 (internal quotation marks omitted).

117. *Id.* at 161 (internal quotation marks omitted).

118. *Id.* at 162.

119. *Id.*

120. *Id.*

“just outcomes.”¹²¹ Judges also acknowledged a tradeoff between promptness in issuing opinions and opinion quality.¹²² Although Klein did not directly reexamine the Howard categories of judicial types, his survey demonstrated that judges approach their roles very differently. These different approaches are distinct from quality or ability, but they may relate to those attributes.

Judges are a diverse lot, like any group of humans. They value different things in the course of rendering their decisions. Some may be more ideological in their judgments. Some may be more traditionally legalistic in their decisionmaking. Others may be pragmatic. As the surveys show, different judges view their roles differently. Judges place different emphases on the tradeoff between their roles as “generators of precedents” and resolvers of disputes.¹²³ Judges who emphasize their role as precedent-generators might be expected to issue decisions that differ in terms of the features measured by Choi and Gulati’s quality analysis.

A judge’s type may substantially influence citation counts. In Part I.B.1, we discussed the variables that might create statistical “noise” that could interfere with research based on citations but not inherently bias such research. Here, we recognize that there are some systematic biases to the citation-counting system, previously unrecognized, that could bias the results and potentially distort our conclusions about judicial quality. The frequency of citations may capture different perceptions of the judicial role as much as they capture different degrees of judicial quality. This is fine only insofar as the judicial role correlated with higher citations is a normatively desirable role for judges. But the latter claim has not been supported.

The first role of a judge is to decide cases and reach the correct outcome, regardless of the opinion, and the citation studies fail to directly capture this role. Circuit court judges are also expected to write opinions in some cases, which creates the public good of defining the law. The citation studies may capture this measure, but more citations do not necessarily imply better opinions, because this depends on the definition of “better.” One source suggests that an

121. KLEIN, *supra* note 5, at 22 tbl.2.1.

122. *Id.* at 26.

123. See Farber, *supra* note 14, at 1178. He suggests that a “judge who is willing to cut corners on fidelity to the record can publish more opinions,” but notes that such a judge “is also decreasing the ability of the legal system to respond accurately to the facts of cases” and degrading the overall performance of the system and fairness to individual litigants. *Id.*

opinion is cited “as long as it helps to resolve cases.”¹²⁴ More citations could therefore be a useful measure of helpfulness. But it may also be that decisions that are less restrictive and that place greater trust and discretion in the judges deciding future cases will be cited less. If one believed that decisions permitting such discretion were good ones, the citation-counting results could be underinclusive.¹²⁵

Especially clear decisions might also produce fewer citations because of litigant case selection effects. The established Priest-Klein hypothesis suggests that only the difficult cases at the margin go to court, whereas the easy cases are settled.¹²⁶ An opinion that sets out a clear governing rule for future decisions would produce fewer marginal cases than one that establishes a more ambiguous rule. The ambiguous rule would be cited more often because it yields more cases, but it would not necessarily be the preferred rule, nor would its creating judge be the preferred judge.

Other decision characteristics might influence citations but not necessarily be desirable. Simple writing style could affect citations in ways that are not particularly socially valuable.¹²⁷ For example, one might expect the addition of *dicta* to an opinion to increase its citation counts, but this is not necessarily considered a positive addition.¹²⁸ Perhaps longer opinions are more often cited due simply to their length, which is also debatable as a virtue of opinion writing.¹²⁹ Unusually pithy or clever language might be more often cited. At the

124. Montgomery N. Kosma, *Measuring the Influence of Supreme Court Justices*, 27 J. LEGAL STUD. 333, 340 (1998).

125. In a different context, Professor Vladeck suggests that “a less interventionist Court is not necessarily a less effective or influential Court.” Vladeck, *supra* note 22, at 1437.

126. See generally George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. LEGAL STUD. 1 (1984) (proffering empirical explanations as to which cases will reach trial and which will settle).

127. See, e.g., Choi & Gulati, *Mr. Justice Posner?*, *supra* note 1, at 30–31 (observing that “the judge who invests a great deal of resources into writing a good description of the standard of review might obtain a large number of cites” but that those resources might be better spent on other tasks). Professors Choi and Gulati also note that certain judges, such as those with an academic background, may “write in an especially provocative and pedagogic fashion” that might generate high citation numbers independent of the substance of their opinions. *Id.* at 32; see also Vladeck, *supra* note 22, at 1432 (suggesting that the high frequency of citations to Judge Learned Hand was more due to the “nimbleness of his pen” than to “the force of his logic”).

128. See Farber, *supra* note 14, at 1179 (suggesting that citation counts may reflect the “insertion of unnecessary dicta or address issues not raised by the lawyers or addressed by the trial judge,” none of which seem to be “particularly beneficial behaviors”).

129. See David J. Walsh, *On the Meaning and Pattern of Legal Citations: Evidence from State Wrongful Discharge Precedent Cases*, 31 LAW & SOC'Y REV. 337, 357 (1997) (finding that longer opinions received more citations).

circuit court level, reversals of the district court tend to be cited more often than affirmances,¹³⁰ but less deference to district courts may not be desirable. Choice of issues may influence citation counts. A bold judge who decides cases on constitutional grounds might receive more citations than one who narrowly interprets a statute to avoid the constitutional issue.¹³¹

The invocation standard, which measures mentions of judges by name, does not suffer directly from the shortcomings of the citation measure, but it has its own biases. The standard is very likely biased by extrajudicial activities. Judge Posner, for example, has produced an amazing amount of scholarship over and above his judicial effort,¹³² which substantially raises his profile and surely makes it more likely that his name will be invoked in an opinion for its persuasive effect. This might be called a “superstar” effect, as evidenced by the relatively few high outliers on these measures, and it very likely influences citation counts as well.¹³³ Such citations may reflect “familiarity” more than “quality.”¹³⁴

Judges’ qualifications on the traditional citation count or invocation measures are not likely to much affect our Supreme Court review measures discussed above, which are based on decisional outcomes and substantive precedent. One can dispute whether citations or Supreme Court outcomes is the better measure, but that dispute would miss the point that they are measuring two different types of quality. Indeed, the respective qualities may be inverse; Judge Posner suggests that judicial creativity (which might be expected to yield more citations) is likely to increase a judge’s

130. See Brian N. Lizotte, *Publish or Perish: The Electronic Availability of Summary Judgments by Eight District Courts*, 2007 WIS. L. REV. 107, 140 n.159 (arguing that this makes intuitive sense “because a reversal alters the status quo set by the district court, meriting a more careful explanation, and also indicates that the law is uncertain”).

131. There is a canon of statutory interpretation that statutes should be construed to avoid constitutional issues when possible. See, e.g., *NLRB v. Catholic Bishops of Chi.*, 440 U.S. 490, 504–07 (1979) (construing a statute narrowly to avoid a potential constitutional question).

132. Choi & Gulati, *Mr. Justice Posner?*, *supra* note 1, at 24 (observing that Judge Posner “writes multiple books and articles every year”). The same is true for Judge Easterbrook, though to a lesser degree. *Id.* at 25 n.18.

133. See Farber, *supra* note 14, at 1183 (discussing the superstar effect and how it influences citations and invocations). Professor Solum provides a slightly different explanation for this effect. He argues that some decisions are situated more prominently in the network of citations, which causes them to be more frequently cited, which in turn creates a positive feedback loop that induces still more citations. Solum, *supra* note 2 (manuscript at 16).

134. Vladeck, *supra* note 22, at 1432.

reversal rate, as compared to the “unadventurous judge.”¹³⁵ One cannot evaluate the measures themselves without carefully considering what type of judiciary is desired.

This Section will discuss two judging styles, which we call the judicial entrepreneur and the judicial minimalist. These two styles are not the only possible judicial approaches as neither precisely identifies the “unadventurous” career judge bureaucrat type or the leisure-seeking type.¹³⁶ Judicial entrepreneurs and judicial minimalists are not even wholly exclusive categories, but they set out a distinct contrast of approaches to judging that best enables a framing of different theories of the good judge. Judicial entrepreneurs are like Howard’s judicial “innovators,” whereas judicial minimalists are more like his judicial “interpreters.” We shift the semantics to correspond to words more commonly employed in the political science and legal research.

1. *Judicial Entrepreneurs.* The political science literature has addressed a concept it calls “policy entrepreneurs.”¹³⁷ Policy entrepreneurs are individuals who push a new approach to public policy by setting an agenda for change. They are individuals “who specialize in identifying problems and finding solutions.”¹³⁸ The research has focused generally on the legislative process, in which such entrepreneurs are prominent. In one study of twenty-three policy formulation processes, such entrepreneurs were important in fifteen and unimportant in only three cases.¹³⁹ The activities taken by such entrepreneurs include defining and reframing problems, specifying alternatives and advocating ideas, and “exposing such ideas to relevant constituencies.”¹⁴⁰ After identifying some problem, the entrepreneur “must develop strategies for presenting their ideas to

135. Posner, *supra* note 4, at 1277.

136. *See id.* at 1263–64 (discussing unadventurous career, bureaucrat, and leisure-seeking types of judges).

137. *See, e.g.*, FRANK R. BAUMGARTNER & BRYAN D. JONES, AGENDAS AND INSTABILITY IN AMERICAN POLITICS 3–4 (1993) (discussing the concept of policy entrepreneurship).

138. NATHAN W. POLSBY, POLITICAL INNOVATION IN AMERICA: THE POLITICS OF POLICY INITIATION 171 (1984). They are those “willing to invest their resources in return for future policies they favor.” JOHN W. KINGDON, AGENDAS, ALTERNATIVES, AND PUBLIC POLICIES 214 (1984).

139. KINGDON, *supra* note 138, at 189.

140. *See* Nancy C. Roberts & Paula J. King, *Policy Entrepreneurs: Their Activity Structure and Function in the Policy Process*, 2 J. PUB. ADMIN. RES. & THEORY 147, 148 (1991) (setting forth these activities).

others.”¹⁴¹ They must carefully craft their arguments to be persuasive to the relevant audience for their ideas.¹⁴² If successful, their policy innovation will take hold.

Building on this background from politics, one type of circuit court judge might be styled the “judicial entrepreneur.”¹⁴³ The judiciary differs from the political branches, but, like politicians, judges identify legal problems in cases and may attempt to style a new legal approach to such problems. Such an entrepreneur may be a judge who “is alert to the opportunity for innovation, who is willing to invest the resources and assume the risks necessary to offer and develop a genuinely unique legal concept, and who must strategically employ the written word to undertake change.”¹⁴⁴ A judicial entrepreneur exhibits a certain “swashbuckling flair” in decisionmaking.¹⁴⁵ Professor Howard interviewed a judge who proclaimed that the “Courts of Appeals should take a definite lead in innovating in the law—even at the risk of being overruled.”¹⁴⁶ This judge sounds like a judicial entrepreneur.

Prestige is important to one’s ability to be an effective judicial entrepreneur, and it is maintained through “good use of the opinion-writing process.”¹⁴⁷ A judge may maintain prestige by writing a dissent or separate concurrence that serves as a trial balloon or test-markets an idea.¹⁴⁸ The judge must not only write the opinion persuasively, but must make use of conventional materials to avoid being “viewed as reckless.”¹⁴⁹ Like any entrepreneur, to be successful the judge must have a product that appeals to the relevant market and must be able to effectively sell that product. The development of First Amendment

141. Michael Mintrom, *Policy Entrepreneurs and the Diffusion of Innovation*, 41 AM. J. POL. SCI. 738, 739 (1997).

142. *Id.* at 740.

143. For descriptions of the idiosyncratic behavior of such judges, see generally WAYNE V. MCINTOSH & CYNTHIA L. CATES, *JUDICIAL ENTREPRENEURSHIP: THE ROLE OF THE JUDGES IN THE MARKETPLACE OF IDEAS* (1997); Berger & George, *supra* note 88; Cynthia L. Cates & Wayne V. McIntosh, *Retail Jurisprudence: The Judge as Entrepreneur in the Marketplace of Ideas*, 11 J.L. & POL. 709 (1995).

144. MCINTOSH & CATES, *supra* note 143, at 5 (emphasis omitted).

145. See Cates & McIntosh, *supra* note 143, at 710 (ascribing this quality to Justice Oliver Wendell Holmes).

146. HOWARD, *supra* note 115, at 161.

147. Cates & McIntosh, *supra* note 143, at 716.

148. *Id.*

149. *Id.* at 717.

law has been cited as an example of judicial entrepreneurship.¹⁵⁰ Judicial entrepreneurship need not be ideological or policy oriented but might be aimed at advancing a particular interpretive legal model, for example, originalism.

Not all participants in the political or judicial process are entrepreneurs. Professor Howard's interviews did not reveal a widespread commitment to judicial innovation. Some judges are more passive and give less importance to the broader practical implications of their decisions. A minority of judges who are more aggressive in the pursuit of their ideas for the legal system's operation, though, might be considered judicial entrepreneurs. Perhaps judicial entrepreneurship represents a continuum, with different judges having different degrees of entrepreneurial spirit on different legal issues.

There is reason to think that the Choi and Gulati results capture an inclination toward judicial entrepreneurship. One might expect such judges to be more frequently cited because they strive to make the law. Judicial entrepreneurs would also be expected to have more of their cases reviewed by the Supreme Court and, if they are good entrepreneurs, would likely have more cases affirmed. These hypotheses are generally consistent with the results found in Table 7. Thus, the Choi and Gulati quality measures may implicitly presume that judicial entrepreneurship is associated with judicial quality, a presumption they do not support.

Professor Solum effectively makes this argument. His analysis of networks suggests that citation counts are "likely to be a function of originality" in opinion writing.¹⁵¹ Yet he argues that this originality is more of a vice than a virtue, suggesting that the "very best judges" build on prior decisions and are "experts at avoiding originality," whereas the "very worst judges may be the most original."¹⁵² He suggests that it is these bad judges who create originality by using cases "as the vehicles for changing the law, transforming the rules laid down into the rules that the judges prefer."¹⁵³ Although this criticism seems too strong insofar as it could declare a ruling such as *Brown v. Board of Education*¹⁵⁴ to be a uniquely bad decision,¹⁵⁵ the underlying

150. *Id.* at 716.

151. Solum, *supra* note 2 (manuscript at 17).

152. *Id.*

153. *Id.*

154. *Brown v. Bd. of Educ.*, 347 U.S. 483 (1954).

point is important. A judge may certainly be “too” original in his rulings, insufficiently bound to the rule of law.

The judicial entrepreneur type makes a contribution to the law, but the type also has its shortcomings. It embraces a form of judicial activism that is not uniformly applauded. Moreover, the effort at judicial entrepreneurship may undermine other widely held values of judging. In “seeking to increase their influence, judges may reach for broader holdings by ignoring the factual nuances of specific cases” or ignoring the factual record entirely.¹⁵⁶ Thus, judicial entrepreneurship is not an unalloyed good, and the relative value of such judges may be disputed.

2. *Judicial Minimalists.* The notion of judicial minimalism has been popularized, though not coined, by Professor Cass Sunstein. He focuses on the Supreme Court rather than on circuit courts.¹⁵⁷ Sunstein’s criteria for his minimalism include:

that courts should not decide issues unnecessary to the resolution of a case; that courts should refuse to hear cases that are not “ripe” for decision; that courts should avoid deciding constitutional questions; that courts should respect their own precedents; that courts should not issue advisory opinions; that courts should follow prior holdings but not necessarily prior dicta; that courts should exercise the “passive virtues” associated with maintaining silence on great issues of the day.¹⁵⁸

Sunstein’s judicial minimalism has two distinct, though related, aspects: width and depth.¹⁵⁹ The width of a decision is a function of the breadth of the decision rule propounded in the opinion and the degree to which it covers many other situations and possible future cases. The depth of a decision is its philosophical foundation and the degree to which it relies on underlying fundamental principles. Minimalists prefer narrower opinions that don’t have application well beyond the case facts and shallower opinions that are grounded in a more pragmatic foundation than in a philosophical one.

155. See, e.g., Michael C. Dorf, *Identity Politics and the Second Amendment*, 73 FORDHAM L. REV. 549, 556 (2004) (“It is by now commonplace that *Brown* was more a feature of Cold War ideological struggle than of discerning the Equal Protection Clause’s ‘true’ meaning.”).

156. Farber, *supra* note 14, at 1179.

157. See generally CASS R. SUNSTEIN, ONE CASE AT A TIME: JUDICIAL MINIMALISM ON THE SUPREME COURT (1999) (propounding the notion of judicial minimalism).

158. *Id.* at 4–5.

159. *Id.* at 10–14.

Professor Sunstein focuses on minimalist decisions as “democracy-promoting” because they allow the democratically elected branches to resolve controversial issues.¹⁶⁰ A second and typically overlooked effect of minimalism, however, is that it leaves additional discretion to future judges to resolve issues. By eschewing the creation of broad and clear binding rules, the judiciary allows more room for judges to adapt a decision’s principles to the particular facts of their cases. A broad doctrine not only forecloses future legislative judgment but also forecloses the discretion of future judges in making decisions.

Taking an economic turn, Professor Sunstein contends that good judges would “try to minimize the sum of decision costs and error costs.”¹⁶¹ Decision costs are not strictly those associated with deciding the case, but also include those involved in writing an opinion that sets forth the law. Given the limited information available to the judiciary, writing an expansive opinion that seeks to “cover all imaginable situations” could be time consuming and difficult.¹⁶² Such a broad decision is also likely to increase the error costs of the law because unforeseen circumstances will inevitably develop. Sunstein suggests that a “slower and more evolutionary approach, involving the accretion of case-by-case judgments, could produce fewer mistakes on balance, since each decision would be appropriately informed by an understanding of particular facts.”¹⁶³ He cites studies testing the ability of people to undertake comprehensive social engineering and notes that most efforts “produce calamities.”¹⁶⁴

Professor Sunstein characterizes a minimalist approach to judicial decisionmaking as societally beneficial. He does not universally embrace this approach, suggesting that “maximalist” decisionmaking is sometimes appropriate. For our purposes, the salient facts are (a) the distinction between the minimalist and maximalist approaches and (b) the theoretical possibility that a minimalist approach might prove superior. The results of the quantitative research must be viewed in light of these facts.

Judges are aware of different judicial styles, and some may prefer the minimalist style to that of the judicial entrepreneur.

160. *Id.* at 5.

161. *Id.* at 46.

162. *Id.* at 47.

163. *Id.* at 49.

164. *Id.* at 52.

Professor David Klein interviewed one circuit court judge who expressed his personal preference for opinion style as follows:

If I think an opinion is too preachy, trying to make broad law, I get a bit leery. I like the case-by-case approach. With the broad approach, the rule can fail when you encounter something unforeseen.¹⁶⁵

This opinion roughly traces Sunstein's case for judicial minimalism.

The conventional measures of judicial evaluation that we have discussed appear to reflect an implicit preference for the judicial entrepreneur-style of judging. One feature of Professor Sunstein's minimalism is the "constructive use of silence."¹⁶⁶ Silence is seldom cited. Judge Posner observed that the Choi and Gulati criteria "implicitly treat[] judicial creativity as a desirable characteristic of circuit judges" but cautioned that "[n]ot everyone will agree."¹⁶⁷

Not everyone agrees that the judicial minimalist is the optimal judge type, either. Professor Sunstein's proposal for judging has been criticized on various grounds. For some, it leaves too many questions unanswered.¹⁶⁸ One commentator suggests that for a true minimalist, "the Supreme Court would never lay down an intelligible rule to govern another case."¹⁶⁹ For others, it ignores the significance of the law and overemphasizes democratic voting.¹⁷⁰ Sunstein himself is cautious about praising minimalism, suggesting that "it would be foolish to suggest . . . that minimalism is generally a good strategy" because "[e]verything depends on contextual considerations."¹⁷¹

If we cannot find optimal judges who know precisely when to employ minimalism, it is conceivable that the optimal court would

165. KLEIN, *supra* note 5, at 99.

166. SUNSTEIN, *supra* note 157, at 5; *see also* Cass R. Sunstein, *The Supreme Court, 1995 Term—Foreword: Leaving Things Undecided*, 110 HARV. L. REV. 4, 7 (1996) (discussing the "constructive uses of silence" (emphasis omitted)). This is not limited to the linguistic content of opinions, but it extends to the basis for judicial decisions. For example, he suggested that *Bowers v. Hardwick* should have been dismissed for mootness. *Id.* at 68 n.309.

167. Posner, *supra* note 4, at 1276.

168. Neil Devins, *The Democracy-Forcing Constitution*, 97 MICH. L. REV. 1971, 1993 (1999) (book review).

169. Sheldon Gelman, *The Hedgehog, the Fox, and the Minimalist*, 89 GEO. L.J. 2297, 2314 (2001) (book review).

170. *See* Christopher J. Peters, *Assessing the New Judicial Minimalism*, 100 COLUM. L. REV. 1454, 1458 (2000) (arguing that minimalism "significantly underestimates both the legitimacy and the competence of the judiciary in making decisions about individual and minority rights").

171. SUNSTEIN, *supra* note 157, at 50.

contain a blend of judicial types—some judicial entrepreneurs, some judicial minimalists, and some judges with other characteristics.¹⁷² A court consisting entirely of judicial entrepreneurs might be unduly activist and, to the extent that their judicial preferences differ, might produce many conflicting precedents and instability in the law. At the Supreme Court level, we might see cases with nine different opinions and no defined majority ruling. A court of only judicial minimalists, though, might leave the law stagnant and unable to respond to changing societal circumstances.¹⁷³

Our analysis of judicial entrepreneurs, judicial minimalists, and other judicial types is purely theoretical. It is not yet certain whether such types truly exist or how to identify them in the judiciary. Judges' candid answers in interviews are inevitably anonymous. Choi and Gulati and others have collected data that offer some insight into judicial types for particular judges. We can group judges into particular judicial types, at which point each group or type can be analyzed.

B. Cluster Analysis of Judges and Types

Cluster analysis is a statistical tool that enables like items to be grouped based on similarities and dissimilarities across a set of variables. This procedure has been called the practice of “finding groups in data.”¹⁷⁴ The procedure identifies the cases (individual judges) who share similarities with respect to certain variables (for example, citations and *Reviewscore*) and groups them into categories, creating a sort of taxonomy of judges. The method also allows for a measure of the distance between separate clusters. Although this technique does not directly measure judicial type, such as entrepreneur or minimalist, it does enable us to identify judges who apparently share characteristics and evaluate their common features. Cluster analysis is frequently used before hypotheses are formed, in the exploratory phase of research, as is the case here. For example,

172. The effect of different judge types on precedent has been modeled. Sophie Harnay & Alain Marciano, *Judicial Conformity Versus Dissidence: An Economic Analysis of Judicial Precedent*, 23 INT'L REV. L. & ECON. 405, 411–17 (2003).

173. On the need for and benefits of some judicial creativity and problem solving, see generally Frank B. Cross, *What Do Judges Want?*, 87 TEX. L. REV. 183 (2008) (book review).

174. LEONARD KAUFMAN & PETER J. ROUSSEEUW, FINDING GROUPS IN DATA: AN INTRODUCTION TO CLUSTER ANALYSIS 1 (2005). For a further general review of the procedure, see generally BRIAN S. EVERITT, SABINE LANDAU & MORVEN LEESE, CLUSTER ANALYSIS (4th ed. 2001).

the method is valued in the field of psychiatry, in which the “correct diagnosis of clusters of symptoms such as paranoia, schizophrenia, etc. is essential for successful therapy.”¹⁷⁵

A consequence of any quality-ranking system for judges is that it considers only one dimension of quality or depends on the calculator’s valuations in comparatively weighting multiple dimensions.¹⁷⁶ There are no indubitable weights to be ascribed to particular measurable judicial qualities; commentators have not even agreed on the qualities, much less their measures or weights. Hence, we do not propound a list of the “best” circuit court judges, which could at best reflect only our own personal valuation of judicial types. Instead, we set forth the different types of judges, so that users may draw their own conclusions about who is best or who should be promoted to the Supreme Court.

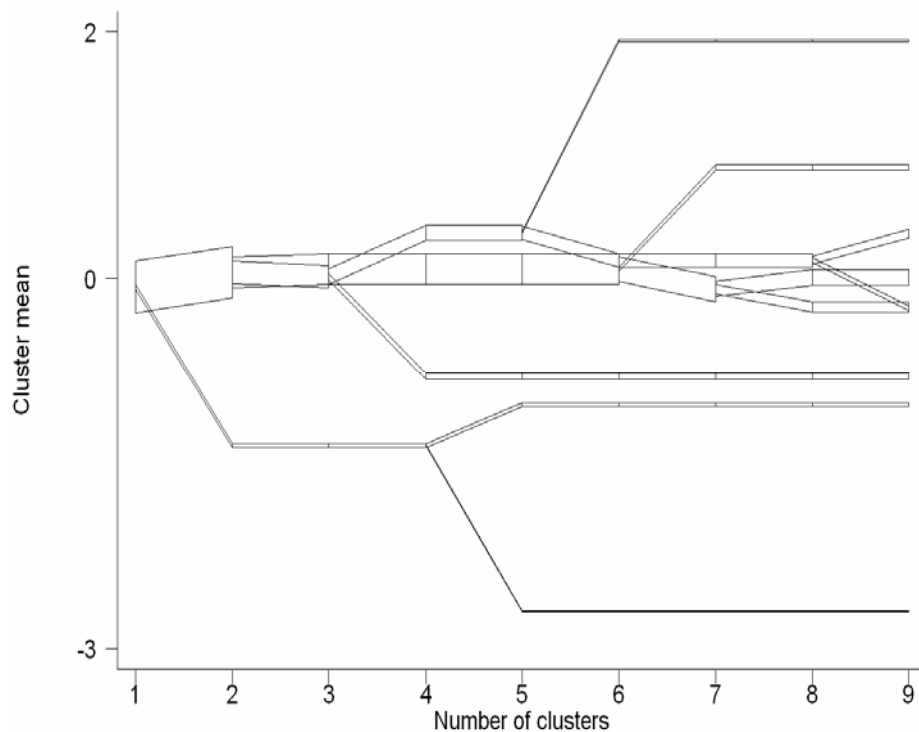
For our cluster analysis to define types of judges, we used the Choi and Gulati quality variable of citations, *Reviewscore*, and the judge’s Almanac evaluations from attorneys. The use of these variables, covering different time periods, limited the sample of judges measured to fifty-eight.¹⁷⁷ The procedure requires that a certain number of clusters be chosen, and we selected nine clusters. Figure 2 graphically displays the cluster results in a dendrogram.

175. THOMAS HILL & PAWEL LEWICKI, STATISTICS: METHODS AND APPLICATIONS 117 (2006).

176. Judge Posner warns of this effect. Posner, *supra* note 4, at 1276 (“[N]umerical rankings are questionable when the rankings are multidimensional, so that the weighting of the different dimensions becomes critical to the rankings.”). Professors Choi and Gulati are responsive to this concern, though, by calculating the best judges under different weightings ascribed to their different dimensions.

177. The clustering procedure used the Ward’s Algorithm with the variables standardized. It is based on a program described in Matthias Schonlau, *The Clustergram: A Graph for Visualizing Hierarchical and Nonhierarchical Cluster Analyses*, 2 STATA J. 391, 391–402 (2002).

Figure 2. Cluster Analysis



The cluster analysis is merely a description of commonalities and does not directly say anything about the composition of particular clusters. Thus, it does not directly say which cluster might represent judicial entrepreneurs as opposed to judicial minimalists, though we can infer this information from the results. The procedure simply tells us that certain identified judges share commonalities on the dimensions measured. The resulting clusters are as follows:

Cluster 1: Judges Batchelder (6th), Sentelle (D.C.), Seymour (10th), Wilkins (4th), Dubina (11th), Rovner (7th), Henderson (11th), Edmondson (11th), Davis (5th), Scirica (3d), Barksdale (5th), Alito (3d), Loken (8th), Sloviter (3d), Arnold (8th), Boudin (1st)

Cluster 2: Judges Roth (3d), Rymer (9th), Randolph (D.C.), Jones (5th), Kleinfeld (9th), Edwards (D.C.), Ginsburg (D.C.), Martin (6th), Trott (9th)

Cluster 3: Judges King (5th), Niemeyer (4th), Higginbotham (5th), Boggs (6th), Kozinski (9th)

Cluster 4: Judges Smith (5th), Carnes (11th), Ebel (10th), Wilkinson (4th), Bowman (8th), Flaum (7th), Kanne (7th), Wood (7th), Walker (2d)

Cluster 5: Judges Selya (1st), Ripple (7th), Tacha (10th), Jolly (5th), Tjoflat (11th)

Cluster 6: Judges Easterbrook (7th) and Posner (7th)

Cluster 7: Judges Coffey (7th), Torruella (1st), Widener (4th), Demoss (5th), Manion (7th), Garza (5th), Nygaard (3d)

Cluster 8: Judges Pregerson (9th), O'Scannlain (9th), Schroeder (9th), Nelson (9th)

Cluster 9: Judge Reinhardt (9th)

Circuit court judges who share a cluster are more alike on our dimensions of quality than they are like judges from other clusters. These results confirm the Posner/Easterbrook difference suggested by Professors Choi and Gulati. The Posner/Easterbrook cluster is characterized by high citations, high Almanac ratings, and relatively high *Reviewscore* measures.

Justice Alito appears in the first cluster, which is also the largest and one of the more distinct clusters. This cluster is characterized by moderately high Almanac ratings and *Reviewscores* and relatively low Choi and Gulati quality counts. Justice Alito and fellow members of his cluster are quite different from Judges Posner and Easterbrook on our measure. None of our cluster measures made any attempt to capture the judges' ideological orientations.

Whether Judges Posner or Easterbrook or Justice Alito are the optimal choices for the Supreme Court is a separate question. Intuitively, it appears that the clusters may break down along the minimalism/entrepreneurialism distinction: the first cluster includes those of a more judicial minimalist bent, whereas the bottom clusters include those who tend to behave like judicial entrepreneurs (as reflected in higher citations and more frequent Supreme Court

review).¹⁷⁸ Under this theory, Justice Alito was a minimalist selection. This is not determinate, though, absent an objective test for minimalism. Now that we have clustered some of the circuit court judges, though, it becomes possible to conduct further research on the decisionmaking characteristics of particular clusters, to evaluate them for promotion or otherwise. Further research might introduce additional variables that revise the clusters. Significant additional research on circuit court judges is discussed in Part IV.

IV. IMPORTANT ADDITIONAL RESEARCH

Identifying desirable judges is only the beginning of necessary empirical evaluative research on the judiciary. If researchers cannot identify particular standards for judging or particular quality judges, vital questions will go unanswered. Conversely, if researchers can identify those factors, they can shed considerable light on the judiciary. This understanding may enable the selection of better judges. It will also be valuable to see how patterns of decisionmaking vary between better judges and those deemed to be of lesser quality.

One important body of research would attempt to identify the characteristics associated with higher-quality circuit court judges. As noted above, the circuit courts, in bulk, are more important to the state of American law than the Supreme Court. After determining the best sort of circuit court judiciary, researchers can try to identify the background features associated with the best circuit court judges. Thus, it may be possible to identify whether diversity, prior district court or professorial experience, prior publications, law school education, or ABA ratings tend to correspond with better-quality circuit court judges.

Professors Landes, Lessig, and Solimine have preliminarily explored this question. Using the influence measure we have described, they found that measures of academic achievement had no significant effects on influence but that graduation from Harvard or Yale Law Schools had a statistically significant positive effect.¹⁷⁹ Race, gender, and prior judicial experience were insignificant factors.¹⁸⁰ ABA ratings were significant with respect to judges classified as

178. Clusters 8 and 9 are most distinctly characterized by very low *Reviewscore* measures, a feature shared by many of the Ninth Circuit judges. Clusters 4 through 9 have higher Choi and Gulati quality scores than the first three clusters.

179. Landes et al., *supra* note 26, at 324.

180. *Id.* at 324–25.

unqualified, but not with respect to other ratings levels.¹⁸¹ Montgomery Kosma conducted his own study of citations and Justices at the Supreme Court level. He found that the appointment of younger candidates does not increase their influence and that prior service as a private attorney had the biggest positive impact on his measure of judicial value.¹⁸²

Similar research has been conducted on the Australian judiciary.¹⁸³ This study used as measures of judicial “prestige” the total number of citations to a judge’s opinions and invocations of the judge’s name.¹⁸⁴ Although the study reveals some association between prior judicial experience and quality, the strongest positive associations found are between prior academic experience and prior experience as a barrister.¹⁸⁵ Thus, to the extent that one embraces these studies’ measure of judicial prestige, one may better screen appointees for quality on the bench. Certain types of backgrounds may associate with a particular cluster of judging characteristics that is deemed desirable.

Other important research would look at associations between judicial types or objective measures of judicial quality. We have discussed the theoretical arguments for and against judicial minimalism. Empirical evidence could improve these arguments. Once clusters of judge types have been identified, they may be evaluated on scales other than citations or other unidimensional standards. Perhaps one set of characteristics produces better, more pragmatic societal results than do other sets. These topics have been the subject of much musing, but the existence of quantitative data enables more rigorous testing of the degree to which particular judicial types are likely to produce particular results. The data further allow evaluation of whether those results are positive or negative.

The role of judicial ideology is a ripe area for such study. The virtue of ideological decisionmaking is contested. One group, including those like Peretti, embraces decisions grounded in judicial ideology. Another, probably larger, group believes that judges should

181. *Id.* at 325.

182. Kosma, *supra* note 124, at 368–70 & n.70.

183. Russell Smyth & Mita Bhattacharya, *What Determines Judicial Prestige? An Empirical Analysis for Judges of the Federal Court of Australia*, 5 AM. L. & ECON. REV. 233, 241–59 (2003).

184. *Id.* at 241–42.

185. *Id.* at 255–56.

minimize the role of ideology in their decisions insofar as possible.¹⁸⁶ Still others might prefer a moderate role for ideology. Regardless of one's normative opinions, there is value in identifying the judicial characteristics associated with ideological decisions, to determine which characteristics merit approbation or disapprobation. Intuitively, one might expect judicial entrepreneurs or innovators to be the most ideological.¹⁸⁷ This is not necessarily the case, though.¹⁸⁸ The empirical data and cluster analysis enable researchers to test this hypothesis for accuracy.

Researchers could examine the clusters of judges to see if a particular set tended to be more or less ideological in decisionmaking. The basic measures of citation influence or prestige also could be used to see if the more influential judges tend to be more or less ideological. If this is true, one might suspect that different judicial types serve as little more than a mask for a desire to render ideologically friendly decisions. As is often the case, Professor J. Woodford Howard was ahead of his time and actually examined this question twenty-five years ago, finding that innovators tended to make decisions reflecting "libertarian activism."¹⁸⁹ Howard found a "strong link between the political orientations and role perceptions of these judges."¹⁹⁰ He had a very small and somewhat dated sample, though, and his innovative research design has been neglected in the years since he published his seminal book.

Given the data available in 2009, it is increasingly possible to test hypotheses such as that of Professor Howard. To perform this test, one only needs to identify a judicial type (such as a cluster identified in our analysis or some other measure), calculate the frequency of ideological decisionmaking by the judges, and compare those two measures. Obtaining data on the ideological direction of the judge's decision can be a time-consuming task, but a very preliminary

186. This group presents the classic "countermajoritarian" argument against ideological judicial action. Judges are not elected and are given life tenure so that they cannot be removed if their decisions are contrary to the public will.

187. Professor Howard suggests that judges' ideological orientations are mediated by their perceptions of the policy role they should be playing in determining the policy outcomes of judicial decisions. HOWARD, *supra* note 115, at 172–78.

188. See, e.g., Richard A. Posner, *Against Constitutional Theory*, 73 N.Y.U. L. REV. 1, 9–10 (1998) (discussing disputes over how judicial minimalism should be applied to resolve significant disputes and the influence of "Sunstein's politics").

189. HOWARD, *supra* note 115, at 175.

190. *Id.* at 178. This effect was not universal, though, and it varied by the type of case under consideration. *Id.* at 179.

breakdown is available from the Court of Appeals Database.¹⁹¹ These data enable a measure of the ideology of a judge's votes, and we use the overall mean in the 1980s for Choi and Gulati's top scorers, Judges Easterbrook and Posner; the judge reversed most frequently, Judge Reinhardt; and, for curiosity's sake, Judge Learned Hand. Table 10 reports the percentage of liberal decisions for each of these judges.

Table 10. Ideological Tendencies of Judges

	Percent Liberal
Posner	0.27
Easterbrook	0.37
Reinhardt	0.57
Hand	0.46

These measures conform to casual observer expectations, except that Judge Easterbrook is increasingly considered more conservative than Judge Posner.¹⁹² The specific results should be evaluated with great caution. They involve relatively few cases¹⁹³ at the earliest stage of the judicial careers of every judge but Judge Hand. Moreover, those using the data should control the numbers for factors such as type of case, ideological preferences of colleagues on the panel, procedural posture of the case, and so on. Future research must account for these additional factors. Such research could help identify whether ideological decisionmaking tends to make a judge more or less influential.

Yet another area ripe for testing is the degree to which different types of judges show judicial deference. Deference to district court or agency decisions is a legal command, and deference has been used as a metric for the degree to which judges follow the legal model of

191. The United States Courts of Appeals Database, developed by Professor Donald Songer of the University of South Carolina and others, is available through The Judicial Research Initiative at the University of South Carolina. The Judicial Research Initiative, Dep't of Political Sci., Univ. S.C., Appeals Courts Data, <http://www.cas.sc.edu/poli/juri/appctdata.htm> (last visited Mar. 20, 2009). *See generally* Frank B. Cross, *Comparative Judicial Databases*, 83 JUDICATURE 248, 248-49 (2000) (contrasting this database to others).

192. More recent data with a larger sample size bear this out; Posner is now behaving like a very moderate judge with little ideological leaning. *See* Epstein et al., *supra* note 18, at 320 (scoring Judge Posner near the median in the judicial common space).

193. Judge Easterbrook had the fewest cases of these judges with only forty-six coded in the database.

decisionmaking¹⁹⁴ or are influenced by the preferences of the legislative and executive branches.¹⁹⁵ If one preferred deferential judges, as some legal standards imply, it would be helpful to identify the circuit court judges that were most deferential. One might expect that reversals would receive more citations and that frequently-cited judges or judicial entrepreneurs might be less deferential. A deference analysis could compare a particular type of judge with the frequency with which that judge affirms or reverses lower court opinions. Table 11 reports the percentage of cases in which our test judges affirmed the decision on appeal.

Table 11. Deference Tendencies of Judges

	Deference
Posner	0.65
Easterbrook	0.68
Reinhardt	0.41
Hand	0.64

These results are subject to all of the same qualifications to which the preceding ideology results were subjected. The deference percentages do not suggest that Judges Posner and Easterbrook are especially prone to reverse. Judge Reinhardt, however, has a remarkably high reversal rate in this set of cases. Various other related theories would be good candidates for further testing. Perhaps some judges are more influenced by public opinion in writing their opinions. Criteria might be found for identification of judges that are objectively “bad,” by all of the relevant standards.¹⁹⁶

Additional research might consider the significance of judicial clerks. Clerks bear the brunt of much of the opinion writing for many judges, and the empirical analyses may measure clerks as much as judges. Howard found that most judges view their clerks as “[m]oderately [i]mportant” to their decisions and that some judges regarded their clerks as “[v]ery [i]mportant.”¹⁹⁷ It is possible to

194. See Cross, *supra* note 20, at 1500 (“Given the legal standard’s command of some measure of deference to district courts, one would expect that appellees would prevail more often than appellants if circuit court judges were adhering to the legal model.”).

195. See Frank B. Cross & Blake J. Nelson, *Strategic Institutional Effects on Supreme Court Decisionmaking*, 95 NW. U. L. REV. 1437, 1483–91 (2001) (testing Supreme Court votes based on both legal and institutional deference).

196. See generally Geoffrey P. Miller, *Bad Judges*, 83 TEX. L. REV. 431, 431 (2004) (discussing such “bad judges” and their actions).

197. HOWARD, *supra* note 115, at 151 tbl.5.8.

investigate which judges are considered “best” by clerkship applicants; these judges would presumably attract the highest-quality clerks. One way to measure this information might be to identify the top “feeder” judges for future Supreme Court clerks. From 1994 to 2003, the top ten such judges were as follows.

1. Luttig (4th) (30)
2. Silberman (D.C.) (21)
3. Calabresi (2d) (17)
4. Wilkinson (4th) (17)
5. Kozinski (9th) (16)
6. Tatel (D.C.) (15)
7. Boudin (1st) (14)
8. Edwards (D.C.) (13)
9. D. Ginsburg (D.C.) (10)
10. Williams (D.C.) (10)¹⁹⁸

Judge Posner came next, with nine clerks. There appears to be some preference for D.C. Circuit judges, but not an overwhelming one. By some standard, law students or Justices—or both—regard the D.C. Circuit judges most highly.

Others might empirically study judicial collegiality, which is widely regarded as a positive feature of judging. It is well established that circuit court judges are influenced by other members of the three-judge panel hearing an appeal,¹⁹⁹ and they may also be influenced by the judges of their full circuit who are not present on the panel.²⁰⁰ Howard emphasizes that it is problematic to assume that “votes accurately mirror individual attitudes on collegial courts, where give and take is also expected.”²⁰¹ Consequently, it might be possible to see which judges are most affected by the preferences of other panel members. This could provide a more refined test of individual judge independence than the one that Professors Choi and Gulati offer.

198. “*Feed Me, Stephanie Seymour!*”: *Supreme Court Feeder Judge Rankings*, UNDERNEATH THEIR ROBES, Aug. 17, 2004, http://underneaththeirrobes.blogs.com/main/2004/08/feed_me_seymour.html.

199. See sources cited *supra* note 17.

200. Individual panels face some risk of reversal by their full circuit. See generally Tracey E. George, *The Dynamics and Determinants of the Decision to Grant En Banc Review*, 74 WASH. L. REV. 213 (1999) (discussing *en banc* review).

201. HOWARD, *supra* note 115, at 171.

Researchers might also study litigant behavior. Judges cannot set agendas, which are driven by the litigants who choose to bring the particular cases to the appellate courts. Such litigants might be expected to be strategic in deciding which cases to pursue on appeal.²⁰² A judge's characteristics could be an important part in this strategy; an aggressive litigant might prefer to bring a test case before a judicial entrepreneur, whereas a respondent might prefer to settle such a case. A litigant seeking to change the content of the law would very likely benefit from receiving a favorable opinion from a judge who is likely to yield more citations.

Research at the district court level could also be very helpful. The probability of elevation from the district court to the circuit court level is higher than the probability of elevation from a circuit court to the Supreme Court.²⁰³ Ideology is relatively less important for circuit court appointments. Hence, Professors Choi and Gulati's objective of identifying promotion candidates may be better suited to an evaluation of district court judges than circuit court judges. The criteria could include measures like Choi and Gulati's citation test and circuit court reversal rates. Assessments of the district court judiciary would also have to include trial management ability, and the attorneys' evaluations might be more valuable here. All arguments for the value of measuring circuit court quality, independent of promotion, also counsel for evaluating district court judges. As the ideal district court judges are identified, it may become easier to identify prospective district court judges who best fulfill the desired criteria.

Although the research that has evaluated circuit court judges has immediate value and might be used to assess Supreme Court appointments, its greatest value lies in its potential to spawn additional research on these judges. The vast bulk of empirical research has analyzed the judges only for their presumed ideological interests, as reflected by the appointing president's party or in the more precise point scores we have discussed in this Article. Ideological preferences and personal background do not explain everything about judicial outcomes, though. Moreover, the research

202. For a discussion of this theory, see Cross, *supra* note 20, at 1491–94.

203. See Daniel Klerman, Commentary, *Nonpromotion and Judicial Independence*, 72 S. CAL. L. REV. 455, 461 (1999) (reporting that during the 1990s, the probability of circuit court judge promotion to the Supreme Court was 3 percent and the probability of district court judge promotion to the circuit court was 6 percent).

has generally tested only the ideological outcome of a judge's decision, without any consideration for the content of the opinion and its significance as an influential precedent for future holdings. Studies such as citation analyses provide a starting point for research on the content of opinions, which is far more politically significant than the mere outcome of a case. Thus, this data may be useful to a variety of judicial studies unrelated to Supreme Court appointments.

CONCLUSION

Ultimately, we embrace Professors Choi and Gulati's general research program, which pursues quantitative measurements for judicial characteristics. Like any seminal effort, their approach does not provide the ultimate answer for judicial evaluation, but it offers an excellent starting point. Their tremendous data collection effort will be of great value to future researchers. Their insights also advance the ball considerably for research on relative judicial aptitude. With their data, one can better understand the nature of different judges and draw conclusions about their comparative quality.

The quantitative metrics of the research for circuit court judges do not unambiguously measure the features of judicial quality. Productivity, independence, and citation frequency tell something about the nature of a judge's opinions, but not necessarily whether the judge is particularly meritorious. Given that there are different types of judges with differing concepts of judging, the existing scales may only measure leading judges of a certain type—and that type might not be preferable for the circuit courts or for the U.S. Supreme Court. Determining which particular type may indeed be preferable requires a more extensive theoretical analysis than researchers have conducted. The evaluation of circuit court judges must account for the different theories and practices of judging on the courts.

By combining data sources on circuit court judges, researchers can begin to identify patterns relevant to an assessment of the judges' qualities. One can descriptively differentiate judges based on their judicial types. Those who embrace a Solumonic vision of judicial quality can try to identify the characteristics of judges who fit their model and then better determine which judges possess those characteristics.

Because different judicial types exist, academics cannot objectively identify the best judges. It is important to advance the

discussion beyond obvious judging values that cannot be tested and therefore offer little practical value. Subjective commentators may applaud (or denigrate) a particular judge, but such assessments may simply reflect a bias of the commentator. Academics can provide additional value by presenting quantitative analyses measuring judges' features, which can help policymakers and others ascertain who they believe the best judges are.