SOUTHERN CALIFORNIA
LAW REVIEW

ON TOURNAMENTS FOR APPOINTING GREAT JUSTICES TO THE U.S. SUPREME COURT

WERL

Reprinted from
SOUTHERN CALIFORNIA LAW REVIEW
Volume 78, November 2004, Number 1
© Copyright 2004 by the University of Southern California

UNIVERSITY
OF SOUTHERN
CALIFORNIA

ON TOURNAMENTS FOR APPOINTING GREAT JUSTICES TO THE U.S. SUPREME COURT

WERL*

Professors Stephen Choi and G. Mitu Gulati are ambitious: they seek to quantify great legal minds. This interesting endeavor has both descriptive and normative components. Choi and Gulati hope not only to offer objective measures of what makes a judge great but also to affect the choices that presidents and senators make when appointing individuals to the U.S. Supreme Court. Their procedures—which entail running the names of judges through various judicial tournaments—lead the authors to endorse a handful of well-known federal appellate judges for nomination to the High Court. We admire Choi and Gulati's empirical skills as well as their clever take on an important issue. The ultimate question, however, is whether the current system does any better than the authors' tournament in predicting skills that include the ability to compromise and negotiate, a talent for shaping national policy, and a gift for choosing among the thousands of petitions for certiorari filed with the Court.[1] Choi and Gulati

* "WERL" is the acronym for Washington University's weekly Workshop on Empirical Research in the Law; participants hail from Washington University's law school, political science department, and economics department. The group has been meeting since the Summer of 2000 and we have had countless lively discussions and debates—we think it fair to say that Professors Choi's and Gulati's scholarship (on a variety of topics) have fostered many of these exciting exchanges. For that reason we encourage the authors to continue their intellectual endeavors (notwithstanding our criticisms here). Regular WERL participants, all of whom contributed in some way to these comments, include Kathie Barnes, Randy Calvert, Kathleen Clark, Lee Epstein, Pauline Kim, Andrew Martin, Bob Pollak, Ted Ruger, Margo Schlanger, Nancy Staudt, and Peter Wiedenbeck. We are grateful to the Washington University Law School and the National Science Foundation for supporting our work and to René Lindstädt for providing outstanding research assistance. We used R and Stata to conduct the analyses, and to generate the graphs presented in this article. *See The R Project for Statistical Computing*, *at* http://www.R-project.org (last visited Nov. 16, 2004). The data can be found at the WERL Web site, http://werl.wustl.edu.

1. Stephen Choi & Mitu Gulati, *Choosing the Next Supreme Court Justice: An Empirical Ranking of Judge Performance*, 78 S. CAL. L. REV. 23, 31 n.11 (2004).

say they "do not see how"[2] the current system could do any better than their tournament for identifying the nation's greatest legal minds.

We are not so sure. While we agree it is worthwhile to identify objective measures of legal greatness, we believe the authors replicate—and at times exacerbate—some of the problems associated with the current appointments process. Although Choi and Gulati have undoubtedly started a useful dialogue, we argue that their approach is so fraught with faulty assumptions and imperfect measures that to replace the current nomination procedures with their tournament could lead to a series of harms that may well be more serious than those we face under the existing system. We describe the drawbacks as we see them—with Part I focusing on the troubling assumptions, and Part II on the problematic measures—and note that while the authors could readily fix some of these defects, others are inherent in the tournament and, thus, could not be remedied without rejecting the approach altogether. Consequently, while we applaud the creative thinking that leads the authors to propose a series of judicial games for vetting the names of possible Supreme Court nominees, we ultimately conclude in Part III that presidents and senators should reject Choi and Gulati's reform proposal.

I. PROBLEMATIC ASSUMPTIONS

Embedded within the Choi and Gulati proposal are several occasionally opaque but certainly troubling assumptions. In what follows we consider two: (1) that politics, philosophy, and ideology—and not ethics, competence, and intelligence—now dominate the process by which Supreme Court Justices attain their seats; and (2) that the judges of the federal appellate bench should form the sole pool from which to draw ethical, competent, and intelligent nominees.

A. DOES THE CURRENT PROCESS YIELD UNQUALIFIED JUSTICES?

Academics and policymakers alike have expressed a number of concerns regarding the appointment of Supreme Court Justices.[3] But one rising above nearly all others centers on the increasingly politicized nature

2. *Id.*

3. *See, e.g.*, STEPHEN L. CARTER, THE CONFIRMATION MESS: CLEANING UP THE FEDERAL APPOINTMENTS PROCESS (1994); MARK SILVERSTEIN, JUDICIOUS CHOICES: THE NEW POLITICS OF SUPREME COURT CONFIRMATIONS (updated college ed. 1994); Ruth Bader Ginsburg, *Confirming Supreme Court Justices: Thoughts on the Second Opinion Rendered by the Senate*, 1988 U. ILL. L. REV. 101 (1988).

of the process. Surely it is the case, as Gregory Caldeira and John Wright note, that "the selection of [Supreme Court Justices]—once a 'cozy triangle' of senators, the executive branch, and the bar—[has become] a major arena for the participation of interest groups."[4] It also is true that the media coverage of individual nominations has grown precipitously over the past few decades, such that, for example, the *New York Times* and *Time* magazine ran, on average, about fifty more stories on each post-1980 nominee than they did on his or her predecessor.[5]

These sorts of facts and figures lend some support to the claims of Choi and Gulati, as well as those made by many other observers, that the environment surrounding Senate contemplation of Supreme Court nominees has grown increasingly political and highly charged.[6] But the more relevant question is the extent to which U.S. senators are political in response. Do senators' votes, in other words, attend to ideological and partisan-political concerns rather than to a candidate's qualifications to serve on the High Court?

Certainly Choi and Gulati would answer in the affirmative—indeed, their tournament proposal rests in no small measure on the assumption that politics and not qualifications now dominate the outcomes of Senate confirmation proceedings—and we have uncovered little commentary in the legal literature that would controvert them. But systematic evidence suggests otherwise. In an empirical study we conducted of the twenty-six Supreme Court nominations since 1953, we found that *both* the candidates' ideology (or, more precisely, their ideological distance from the senators) and their qualifications exerted significant effects on the votes cast by individual senators.[7]

4. Gregory A. Caldeira & John R. Wright, *Lobbying for Justice: The Rise of Organized Conflict in the Politics of Federal Judgeships*, *in* CONTEMPLATING COURTS 44 (Lee Epstein ed., 1995).

5. RICHARD DAVIS, ELECTING JUSTICE (forthcoming 2005).

6. Mark Silverstein, for example, observed the following:

> The harsh reality of recent experience—that modern interest groups and media politics shape the selection of judges to our highest courts—has provoked a good deal of concern on the part of politician and citizen alike, and calls for reform of the process of 'advice and consent' are frequently heard.

SILVERSTEIN, *supra* note 3, at 164.

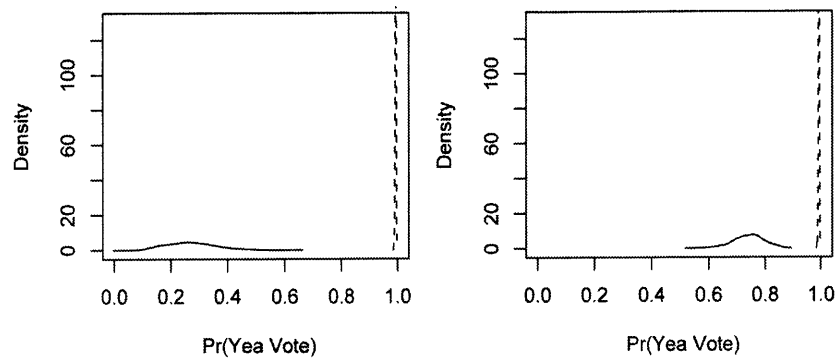
7. For more about our research, see *infra* notes 8–9. For systematic empirical studies reaching the same conclusion, see Charles M. Cameron, Albert D. Cover & Jeffrey A. Segal, *Senate Voting on Supreme Court Nominees: A Neoinstitutional Model*, 84 AM. POL. SCI. REV. 525 (1990); Jeffrey A. Segal, Charles M. Cameron & Albert D. Cover, *A Spatial Model of Roll Call Voting: Senators, Constituents, Presidents, and Interest Groups in Supreme Court Confirmations*, 36 AM. J. POL. SCI. 96 (1992). Our study builds on these.

While the specifics of our study are available elsewhere,[8] Figure 1 conveys the basic flavor of the results. The left panel shows that the probability, across the twenty-six nominees under analysis, of a favorable vote varies depending on a senator's ideological proximity to the candidate under consideration. Keeping in mind that nearly eighty-five percent of all senate votes are "yeas," note that when we set all the other variables in our model at their mean (or median, with the exception of the interaction between ideology and qualification), the likelihood, on average, of a senator voting for a candidate is 23.5% when that senator and the candidate are ideological extremes (the black line). That figure increases to 99.4% when they are at the closest levels (the dashed line).

8. *See* Lee Epstein, Jeffrey A. Segal, Nancy Staudt & René Lindstädt, *The Role of Qualifications in the Confirmation of Nominees to the U.S. Supreme Court*, 32 FLA. ST. U. L. REV. (forthcoming 2005), *available at* http://epstein.wustl.edu/research/qualified.pdf (pre-copy-edited version). *See also infra* note 9. Worth noting here, though, is that we employ Cameron, Cover, and Segal's approach to qualifications: a measure of qualifications derived from a content analysis of newspaper editorials written from the time of nomination by the president until the vote by the Senate. Cameron et al., *supra* note 7, at 529–30. Specifically, they selected four of the nation's leading newspapers, "two with a liberal stance (the *New York Times* and the *Washington Post*) and two with a more conservative outlook (the *Chicago Tribune* and the *Los Angeles Times*)," and identified every editorial that offered an opinion on the candidate's qualifications (from the nominations of Earl Warren through Anthony Kennedy). *Id.* at 529. With the editorials in hand, Cameron and his colleagues coded their content on the basis of claims about the nominee's acceptability from a professional standpoint; the research team then created a scale of qualifications for each nominee that ranged from one (most qualified) to zero (least qualified). *Id.* at 530. In our analysis, we use a nominee's lack of qualifications rather than his or her qualifications as an independent variable. We derive the "Lack of Qualification" variable simply by subtracting the "Qualification" measure from one.

The Cameron et al. approach has at least three advantages for empirical work. First, it has a high degree of facial validity; that is, it appears to comport with our existing knowledge of the nominees. For example, it is Carswell—reckoned "mediocre" even by supporters—who receives the lowest score. (Recall Senator Roman Hruska's (infamous) defense of Judge Carswell: "Even if he were mediocre, there are a lot of mediocre judges and people and lawyers. They are entitled to a little representation, aren't they?" *See* JOHN ANTHONY MALTESE, THE SELLING OF SUPREME COURT NOMINEES 16 (1995)). And it is Kennedy, Ginsburg, Scalia, and several others—candidates even would-be opponents admitted were qualified to serve—who receive the highest. Second, the scores pass standard criteria for intercoder reliability: using π as their index, Cameron and his colleagues report results of 0.87 (p <0.001). Cameron et al., *supra* note 7, at 533. Finally, and perhaps not so stunningly given the range of newspapers consulted, the measure does not appear biased by ideology or political party; in other words, neither Republicans nor Democrats received higher (or lower) ratings based solely on their policy preferences or partisanship. (A bivariate regression of the qualification score on our measure of ideology produces an insignificant coefficient (p=0.136), as does a regression of qualifications on the nominee's political party (p=0.642).)

FIGURE 1. The effect of ideology and qualifications on the votes of senators over Supreme Court nominees, from Justice Earl Warren to Justice Stephen Breyer[9]



Each curve represents the probability density of voting "yea" on the nominee, accounting for the uncertainty in our estimates. All variables are set at their sample means (or medians). (For each panel, the variable on the x-axis is interacted with the sample mean of the other variable.) The left panel shows these probabilities for ideologically distant nominees (black), and ideologically proximate nominees

9. Figure 1 is based on the following data:

Variable	Coefficient	(Std. Err.)
Lack of Qualifications	-0.678**	(0.239)
Ideological Distance	-0.641**	(0.249)
Lack of Qualifications × Ideological Distance	-9.009**	(0.920)
Strong President	0.900**	(0.103)
Same Party	0.367**	(0.089)
Constant	1.507**	(0.094)
N	2461	
Log-likelihood	-584.399	
$\chi^2_{(5)}$	942.23	

This is a probit model, which we estimated using maximum likelihood. The symbol "**" indicates $p \leq$ 0.01. The dependent variable consists of the 2461 confirmation votes cast by individual senators on the nominations of Justice Warren in 1953 through Justice Breyer in 1994. The measures for "Lack of Qualifications," "Strong President," and "Same Party" are the same as those in Cameron et al., *supra* note 7, at 530, and Segal et al., *supra* note 7, at 107. "Ideological Distance" is the ideological distance between a nominee and a senator. For our measurement approach and distance calculations, see the detailed discussion in Epstein et al., *supra* note 8, at 13–22 (further investigating the role of qualifications in the Supreme Court appointment context).

(dashed); the right panel, for highly unqualified nominees (black) and high qualified nominees (dashed).[10]

Nearly as important as the candidate's policy preferences (relative to the senator's), though, is his or her professional merit. To see this, consider the right panel of Figure 1, which shows the probability of a senator voting for a nominee on the basis of the nominee's qualifications. Notice that when a nominee is perceived as highly unqualified (the black line) and all other variables are at their mean, the likelihood of a senator casting a nay vote is twenty-five percent. That probability decreases 125-fold to 0.2% when the nominee is highly qualified (the dashed line).

What these results, along with additional analyses, indicate is that senators will most certainly vote for candidates who are ideologically close to themselves and well-qualified; and they also will almost certainly vote against candidates who are politically distant and not qualified. And, yet, while the odds are high that they will vote for an undeserving candidate who is ideologically proximate (for example, the Southern Democrats and Clement Haynsworth)—thus underscoring the role of politics—it is also the case that they will, under certain conditions, support a candidate whose politics they dislike if they perceive that candidate to be highly meritorious (Republicans and Ruth Bader Ginsburg)—thus underscoring the role of qualifications.

In short, virtually all contemporary writing on the confirmation of Supreme Court nominees has it exactly right: politics play a critical role. Our statistical modeling exercise leaves little doubt that senators are more likely to vote for nominees who share their policy preferences. But that same modeling exercise also leaves little doubt about the role—and the crucial role at that—of qualifications. Whether a candidate is perceived as meritorious also affects senators' votes and, indeed, exerts an effect nearly as strong as ideological proximity. As such, we believe our results serve to undermine an assumption implicit in so many schemes for abrupt change in the appointments process: that merit takes a back seat to politics. Quite the opposite. It appears that qualifications do not now play, nor have they ever played in the contemporary era, a trivial role in the confirmation of Justices.

10. We generated these figures using the program Zelig in R. *See* Gary King, Michael Tomz & Jason Wittenberg, *Making the Most of Statistical Analyses: Improving Interpretation and Presentation*, 44 AM. J. POL. SCI. 347 (2000); Kosuke Imai, Gary King & Olivia Lau, *Zelig: Everyone's Statistical Software, Version 2.0–6*, http://GKing.Harvard.Edu/zelig (last visited Nov. 16, 2004).

B. DO GREAT APPELLATE COURT JUDGES MAKE GREAT SUPREME COURT JUSTICES?

Choi and Gulati assume we should draw names for promotion to the Supreme Court from federal appellate courts.[11] We cannot help but note the irony in this choice given the lack of confidence the authors have in the current appointments process. More seriously, though, we think there are obvious problems with this idea that the authors never investigate. First, selection procedures that permit presidents to appoint and senators to confirm only appellate court judges work to homogenize the Court in a number of ways that we find harmful.

The limited pool proposed by Choi and Gulati assures that the nominees will all have similar career experiences and this, in turn, leads to a narrow set of possible judicial outcomes on important legal questions. This is not mere conjecture; at least twenty-two empirical studies have investigated the impact of judges' background characteristics and close to seventy percent of the studies found a close correlation between judges' career history and judicial decisionmaking.[12] For example, one study found that prior professional experience impacted judicial views on the constitutionality of the Sentencing Guidelines—judges with past experience as criminal defense lawyers were likely to oppose them while judges with prior judicial experience were likely to find the Guidelines constitutionally valid.[13] Another empirical study found that federal appellate judges with experience representing management clients in union cases were significantly less likely to publish their opinions in federal reporters than were other panels that lacked this experience.[14] We believe these results, as well as many others found in the extant literature, underscore the importance of career diversity on the Supreme Court. "Because judges with varied career experiences bring distinct perspectives to the bench—perspectives that ultimately lead them to make distinct judicial choices—merging jurists with diverse career paths [is more likely to] lead to more effective decision making."[15]

11. Choi & Gulati, *supra* note 1, at 31.

12. 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, 954 (2003).

13. 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, 1383 (1998).

14. Deborah Jones Merritt & James J. Brudney, *Stalking Secret Law: What Predicts Publication in the United States Courts of Appeals*, 54 VAND. L. REV. 71, 102–03 (2001).

15. Epstein et al., *supra* note 12, at 956.

Moreover, looking to federal appellate judges for possible promotion to the Supreme Court solidifies another type of homogeneity—lack of racial and gender diversity.[16] We do not recount here why variation along gender and racial or ethnic lines is desirable; scores of scholars, commentators, and policymakers have already done so. We note only that limiting the pool in the manner that Choi and Gulati propose means that presidents will consider only a handful of women and minorities for promotion. Of the 167 federal appellate court judges in active service, only thirty-six are women and just twenty-four are members of racial or ethnic minority groups.[17] These numbers begin to border on the trivial when we consider that presidents typically nominate members of their own political party to the Court. With a wider pool that includes legislatures, state courts, government legal offices, and private law practices, the selection bias exacerbated by Choi and Gulati's tournament approach begins to dissipate as the number of females and minority group members increases in these other venues. This, in turn, will assure the diversity that so many scholars and commentators argue is necessary for good decisionmaking.

The selection bias and the resulting lack of diversity along numerous axes, however, are not the major flaws in Professors Choi and Gulati's article in our view. The authors, inexplicably, fail to theorize fully the question of "what makes a great Supreme Court Justice?," leaving the reader to wonder why they are comfortable limiting their pool of nominees (given the problems described above) and at the same time adopt the particular measure they do (with the problems we describe below[18]). We think the question is an important one, but that the authors fail to offer a convincing answer. They hint that negotiation, prioritization, and conflict resolution skills are key, but ultimately they rely on their own intuition—a great federal appellate court judge makes a great Supreme Court Justice. "After all," they observe, "the job at the lower court level (deciding cases and writing opinions explaining the decisions) is often much the same as at the higher level (the difference being that one hears fewer, but more important, cases)."[19] We wish the authors had made a more serious attempt

16. *See id.* at 956–57.

17. Sheldon Goldman, Elliot Slotnick, Gerard Gryski, Gary Zuk & Sara Schiavoni, *W. Bush Remaking the Judiciary: Like Father Like Son?*, 86 JUDICATURE 282, 294 (2003). The judges listed in Choi and Gulati's study include twenty-two women and thirteen racial or ethnic minorities. We obtained this data by comparing the information found in Choi & Gulati, *supra* note 1, at 165 app. tbl. A, with Gary Zuk, Deborah J. Barrow & Gerard S. Gryski, *The S. Sidney Ulmer Project: Attributes of U.S. Appeals Court Judges, 1801–1994*, *at* http://www.as.uky.edu/polisci/ulmerproject/auburndata.htm (last visited Oct. 12, 2004).

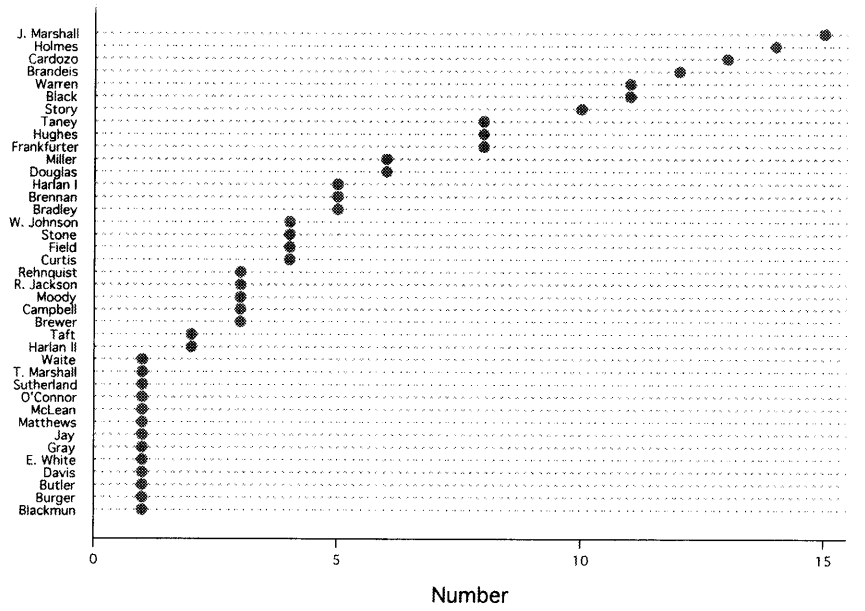
18. See *infra* Part II.

19. Choi & Gulati, *supra* note 1, at 31.

at theorizing greatness at the Supreme Court level, thereby giving insight into their research design and methodological choices.

Although we do not offer a positive theory of the "great justice" here, we did investigate just how past and present Justices would perform in a model that restricts the appointments process to appellate judges. We found sixteen studies in the law and social science literatures ranking Supreme Court Justices according to their perceived greatness.[20] The authors of the studies determined who is great based on their own (or survey respondents') subjective viewpoints and intuitions, and for this reason, they share some of the same drawbacks found in the Choi and Gulati study. But we do find the studies revealing for two reasons. First, the same names show up over and over again. Second, the studies link greatness, however asystematically, to the Justices' actual performance on the Supreme Court. Figure 2 depicts the names of thirty-nine Justices and the number of studies that rated each "great."

FIGURE 2. Justices rated "great" in sixteen studies[21]



"Number" indicates the number of studies rating the justice as "great."

20. *See infra* Appendix.

21. *See infra* Appendix.

The third reason why Figure 2 is revealing is that it highlights the fact that if past presidents had limited their pool of nominees to appellate judges, they would have been limited to but six of the thirty-nine Justices—none of whom received ratings of "great" by more than three experts.[22]

Choi and Gulati might fairly object to the list of names found in Figure 2 given the fact that Congress did not create a separate court of appeals until 1869 and consequently many of the great Justices had no opportunity to serve as federal appellate judges before being nominated to the Supreme Court.[23] If we look only to post-1869 appointments, the list above eliminates the following eleven Justices: Campbell, Curtis, Davis, Field, Jay, W. Johnson, J. Marshall, McLean, Miller, Story, and Taney.[24] While we think this adjustment is necessary, we do note that if presidents desired to appoint federal judges to the Supreme Court prior to 1869 they could have looked to federal district courts. But none—not a single one—of the eleven Justices eliminated from the figure had federal district court experience.[25] Finally, it is worth noting that in the only study limited in scope to Justices of the twentieth century (from Holmes to Breyer), only one appellate court judge (Harlan II) received a rating of "excellent," even though there were plenty of potential candidates by this point in time. Instead, the balance of the list consisted of the familiar greats (all of whom lacked federal appellate court experience): Holmes, Brandeis, Cardozo, Brennan, Warren, Hughes, Black, Stone, Frankfurter, and R. Jackson.[26] In short, if Choi and Gulati are to have their way—with presidents nominating only individuals with federal court experience—then we must forego the wisdom of jurists like John Marshall, Oliver Wendell Holmes, Benjamin

22. *See supra* Figure 2 and *infra* Appendix.

23. The Judiciary Act of 1869 provided for a separate circuit judiciary of nine members. Judiciary Act of 1869, ch. 22, 16 Stat. 44 (1869) (repealed 1911). The establishment of the circuit courts of appeals (renamed the courts of appeals in 1948) did not come until the Judiciary Act of 1891. *See* Evarts Act, ch. 517, 26 Stat. 826 (1891) (codified as amended in scattered sections of 28 U.S.C.). *See also* Epstein et al., *supra* note 12, at 916–17 (discussing percentage of Justices with judgeships when nominated to the Supreme Court).

24. Lee Epstein, Thomas G. Walker, Nancy Staudt, Scott Hendrickson & Jason Roberts, *The U.S. Supreme Court Justices Database* (2004) [hereinafter *Database*], *available at* http://epstein.wustl.edu/research/justicesdata.html (last visited Nov. 16, 2004).

25. *Database*, *supra* note 24. It is worth noting that three Justices appointed prior to 1869 had served on a federal district court: Philip P. Barbour, Peter Daniel, and Robert Trimble. *Id.* None appears on any of the sixteen lists of great Justices. *See supra* Figure 2.

26. Michael Comiskey, Has the Modern Senate Confirmation Process Affected the Quality of U.S. Supreme Court Justices? (unpublished manuscript, on file with authors), *cited in* HENRY J. ABRAHAM, JUSTICES, PRESIDENTS, AND SENATORS: A HISTORY OF THE U.S. SUPREME COURT APPOINTMENTS FROM WASHINGTON TO CLINTON 372 (new & rev. ed. 1999). Interestingly enough, as Professor Abraham notes, Comiskey answers his research question in the negative. ABRAHAM, *supra* at 372.

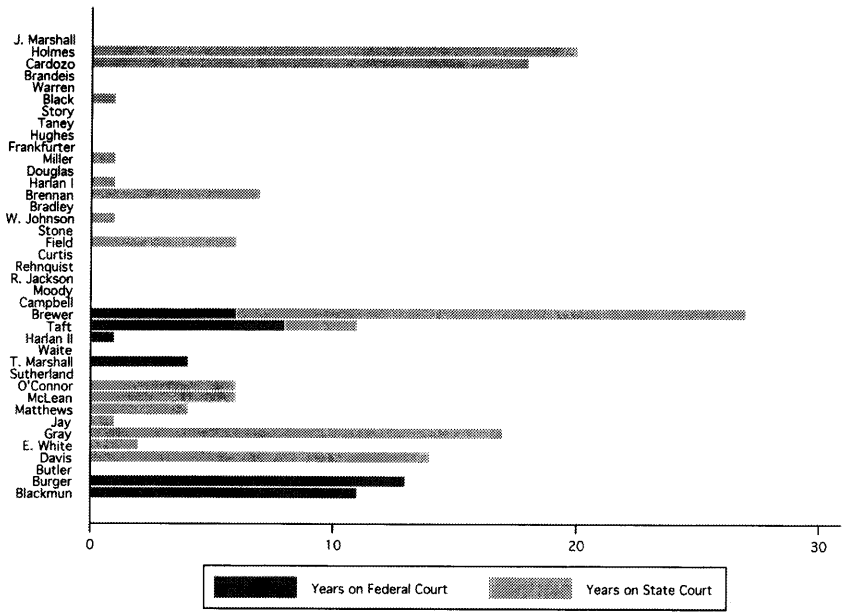
Cardozo, Louis Brandeis, Felix Frankfurter, and Harlan Fiske Stone, to name just a few.

The fact that Choi and Gulati's tournament could not produce Justices widely perceived to be great suggests that we (and they!) should seriously question both the underlying methodology and their empirical findings. We do not think Figure 2 proves that Choi and Gulati's study is fatally flawed (after all, the great Justices were never contestants in judicial games similar to the authors' tournament), but it does mean that further investigation is essential before we accept their scheme as superior to the existing approach to appointments.

Moreover, if judicial experience is in fact relevant, then we wonder if Choi and Gulati looked to the wrong pool. Perhaps they should have made state supreme court experience the prerequisite to a Supreme Court nomination. Like federal appellate court judges, state supreme court justices must decide cases and write opinions explaining their decisions (similarities to which Choi and Gulati point when suggesting federal appellate court judges undertake work similar to that found in the U.S. Supreme Court) but state supreme court justices also tend to sit in groups larger than three, typically must prioritize issues in the case selection process, and arguably have similar conflict resolution problems as those found in the U.S. Supreme Court. Justices Holmes and Cardozo both had prior judicial experience in a state supreme court and nearly all the experts believe they turned out to be great Justices. Indeed, if a prior career as a judge is an important factor for promotion to the High Court, we are not sure why Choi and Gulati did not expand the tournament to all federal and state court judges. As Figure 3 depicts, of the thirty-nine Justices deemed great, six served as federal judges, but seventeen had state court experience (with nine of the seventeen having sat on a state court of last resort).[27] Certainly, we do not claim that these data prove state court experience is more highly correlated to greatness on the Supreme Court, but it does seem to be a factor we would want to consider before eliminating these judges from the judicial tournament.

27. *Database*, *supra* note 24. *See also infra* Figure 4.

FIGURE 3. Justices rated "great" who had judicial experience prior to serving on the U.S. Supreme Court



The bars show the number of years of service.[28] The Justices are ordered on the basis of the number of studies rating them as "great."[29]

Ultimately we question whether any system that relies on past judicial experience as a predictor for legal and judicial greatness is a good system. And we are not alone in this sentiment. Justice Felix Frankfurter, for example, famously quipped, "One is entitled to say without qualification that the correlation between prior judicial experience and fitness for the functions of the Supreme Court is zero."[30] Justice Sherman Minton responded to Frankfurter's comment by suggesting that it be sent to every member of Congress as he thought himself "a living example that judicial experience . . . doesn't make one prescient."[31] We note that while Justice Minton may not have had a great legal mind, he did have an abundance of self-awareness; after all, not a single expert (or student) has ever rated him

28. We derived the data on judicial experience from *Database*, *supra* note 24.

29. *See supra* Figure 2.

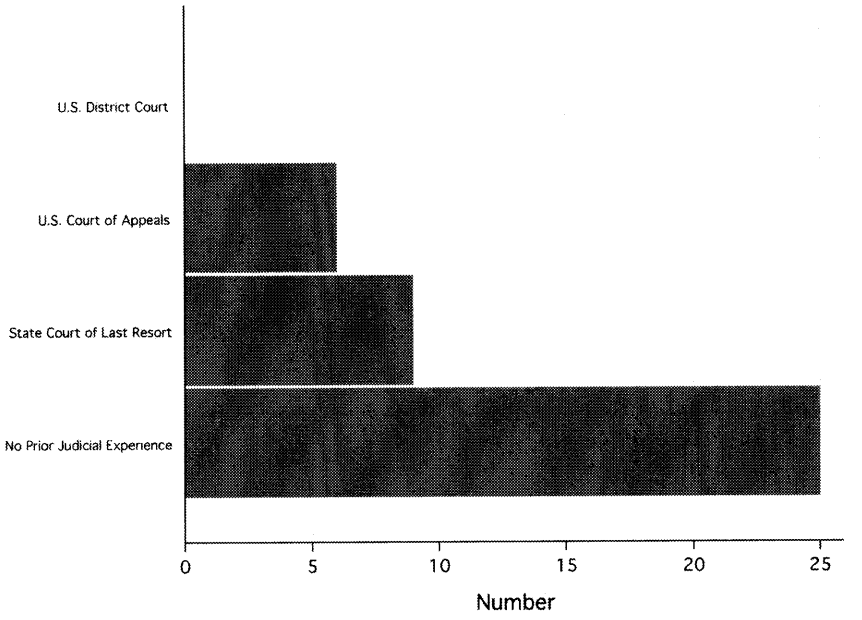
30. Felix Frankfurter, *The Supreme Court in the Mirror of Justices*, 105 U. PA. L. REV. 781, 795 (1957).

31. HENRY J. ABRAHAM, THE JUDICIAL PROCESS: AN INTRODUCTORY ANALYSIS OF THE COURTS OF THE UNITED STATES, ENGLAND, AND FRANCE 61 n.98 (7th ed. 1998) (internal quotations omitted).

to be a great justice, notwithstanding the eight years he served on the Court of Appeals for the Seventh Circuit.

Figure 4 offers some support for the observation that presidents and senators should view prior judicial experience as less relevant than Choi and Gulati desire for nomination and appointment to the Court: of the thirty-nine great Justices, 64.10% (n=25) had no such prior experience. A broader investigation into the past experience of all 148 nominees over the history of the appointments process reveals that the majority lacked the skills and knowledge that Choi and Gulati seem to think is necessary for greatness on the Supreme Court.[32] Also worthy of emphasis is that of the 119 successful nominees, just twenty-one spent time on a federal court of appeals.[33]

FIGURE 4. Number of "great" Justices (n=39) with prior judicial experience.[34]



Professors Choi and Gulati might argue that this lack of experience is precisely what is wrong with the current appointments process: presidents focus more on their ideological litmus tests than on the experience

32. *See Database*, *supra* note 24.

33. *Id.*

34. Justices could be depicted more than once if they served on a federal court *and* a state court of last resort. *See supra* Figure 3.

necessary for greatness. Put differently, perhaps we would have had many more great Justices had the appointments process involved a tournament similar to that proposed by the authors. While we cannot offer definitive evidence proving or disproving this empirical claim—and neither can Choi and Gulati—we do note that this claim is inconsistent with the opinions held by students of the Court. Many commentators argue that by making federal appellate court judges the "darlings"[35] of the appointments process, presidents do not improve the Court, but undermine it.[36] Chief Justice Rehnquist recently observed that at one time his Court housed Justices "drawn from a wide diversity of professional backgrounds," but now no such diversity exists.[37] As the only sitting Supreme Court Justice without past judicial experience, Rehnquist argued that the current appointments process that entails promotion of only judges threatens to produce a Court that resembles those found in the civil law countries—courts comprised of professional jurists who have no respect in the larger population nor independence from the other branches of government.[38] For the reasons outlined above regarding the problems associated with homogeneity, we agree with the Chief Justice and hesitate to applaud any legal reform that entrenches the norm of prior judicial experience rather than disrupts it.

II. PROBLEMATIC MEASURES

The drawbacks and limitations that we identified above lead us to reject Choi and Gulati's suggestion that the pool be narrowly limited to federal appellate judges. Expanding the pool, however, means that the factors to which the authors look for predicting greatness are also problematic given that these factors cannot be easily operationalized in other venues such as the legislatures, government legal offices, or private law practices. In this section, however, we set aside this overarching problem and investigate the factors and measures they chose in the context of a tournament of federal appellate court judges.

Choi and Gulati tell us that their tournament evaluates federal appellate judges based on their productivity, quality of opinion writing, and independence from the views of colleagues and political sponsors. In fact,

35. DAVID ALISTAIR YALOF, PURSUIT OF JUSTICES: PRESIDENTIAL POLITICS AND THE SELECTION OF SUPREME COURT NOMINEES 170 (1999).

36. *See* Epstein et al., *supra* note 12, at 940–41.

37. *See* William H. Rehnquist, *2001 Year-End Report on the Federal Judiciary*, http://www.supremecourtus.gov/publicinfo/year-end/2001year-endreport.html (last updated Jan. 2, 2002).

38. *See id.*

the outcome turns on the number of opinions published by each judge, the number of times his or her opinions are cited by others, and the number of times the judge dissents from opinions by another judge that was appointed by a president of the same political party. A considerable gap thus exists between what the numbers purport to measure and what they actually measure, and this gap is particularly troublesome if the actual measures used are biased in some way. We return to these points below, but first we consider their chosen criteria.

In designing their tournament, the authors attempt to capture "widely-held notions of merit." Perhaps it is uncontroversial that Supreme Court Justices should be productive, write high quality opinions, and exercise independence of thought. But are those qualities all that are required, or are they even the most important qualities that make a Supreme Court Justice great? What about collegiality, the ability to build consensus, and the willingness to engage constructively with a diverse group of colleagues? Perhaps an understanding of the political process or sensitivity to widely held norms are equally or more important in making a Supreme Court Justice great.

These questions suggest that perhaps what makes a Supreme Court Justice great is not some invariant quality across time, but rather the interaction of a number of variables associated with a particular individual at a unique moment in time. Put differently, a great Supreme Court Justice from one era might not be so great sitting on a different Court in a different political era. And, if diversity of experiences makes for a better Court as a whole, as we have suggested above, then perhaps "merit" cannot be measured in a static way. The "best" candidate for a particular opening on the Supreme Court may depend in part upon who else is already on the Court.

We do not propose an alternative list of criteria, we only suggest that without a more complete account of what makes a great Supreme Court Justice, it is not clear why a tournament should turn on the qualities Choi and Gulati have chosen and not others. Of course, the authors have at least one very good reason for basing their tournament on the qualities they have selected—as they themselves admit, they are measures on which data is easy to collect and analyze.[39] By doing so, however, they run the risk of looking for great Supreme Court Justices on the side of the street with the lamppost, rather than wherever they are likely to be found.

39. Choi & Gulati, *supra* note 1, at 29–30.

If, in fact, the number of opinions produced, citation counts, and willingness to take contrary positions together measure greatness, we wonder what exactly they suggest a person is great at doing. In our review of the tournament, we could not help but notice an odd similarity: the same factors that the authors believe make appellate judges particularly qualified for promotion to the Supreme Court also make a great law school tenure file! This resemblance is not by itself problematic, but the fact that former law professors consistently won the tournament games gave us pause. After all, this narrows the pool even more—Choi and Gulati's tournament now tells us that the best Supreme Court Justices will not only be appellate court judges but the ones with past experience in the legal academy.

Consider, for example, the productivity factor, which the authors measure by looking to the number of opinions that each appeals court judge published from 1998 to 2000. We entered their data on the winners of the various tournaments[40] into a database and then added a number of background variables about each judge including career experience, gender, race, and the party of the appointing president.[41] When we compared the productivity variable (adjusted for intercircuit differences[42]) and prior experience as a full-time law professor, a statistically significant relationship emerged—former law professors are more "productive" than their judicial colleagues without prior academic appointment.[43]

In the absence of an underlying theory of greatness, we are not sure how to think about this finding. On the one hand, we hesitate to endorse a measure that privileges fast and furious writing; we believe it possible that in judicial matters, contemplative writing may in the end produce just as good—if not better—decisions. On the other hand, highly productive authors are certainly capable of producing great publications, and many do. Of course, this is the age-old debate heard on university faculties all over the country; should institutions favor slow and methodical authors with a few great pieces or extremely productive authors with a range of contributions that are not all great? Again, without further thought and consideration—and in the absence of any theory of greatness—we do not know how this should be resolved in the nomination process.

40. *See* Choi & Gulati, *supra* note 1, at 83 app. tbls. A–H.
41. These data are available in Zuk et al., *supra* note 17.
42. Choi & Gulati, *supra* note 1, at 90 app. tbl. C.
43. $p = 0.029$ for a Mann-Whitney test of differences in ranks. Note that law professor status is coded 1 if the judge was a full-time law professor and coded 0 if not.

We turn now to the question of whether the data compiled by the authors offer an accurate way of measuring and comparing the qualities they are trying to capture.

A. PRODUCTIVITY

Counting the number of published opinions produced by a judge undoubtedly offers some measure of productivity, reflecting the level of effort expended. Other measures are possible—for example, counting the number of pages produced—but counting the number of published opinions (and excluding unpublished opinions and extrajudicial writings) seems a reasonable choice to us. The difficulty arises when comparing the productivity of judges across circuits. As Choi and Gulati point out, publication rates vary widely between circuits, with the mean number of published opinions per judge ranging from a high of 185.2 (in the Seventh Circuit) to 60.1 (in the Third Circuit).[44] Because these varying rates likely reflect differences in caseload, case complexity, and circuit norms regarding publication, as well as individual differences, the absolute number of published opinions by a judge in one circuit is not directly comparable to that of a judge in another circuit.

Choi and Gulati's solution is to adjust the judges' productivity measures by adding the difference between the mean number of published opinions per judge in their circuit and the Seventh Circuit. They argue that the resulting ranking thus reflects each judge's standing relative to other judges within her own circuit. We question whether this adjustment adequately solves the problem of intercircuit comparisons. Put concretely, does the fact that Judge Reinhardt published eighty-one opinions above the Ninth Circuit average really warrant him a higher productivity ranking than Judge Easterbrook who published only forty-eight above the Seventh Circuit mean, even though Easterbrook in fact published 233 opinions compared to Reinhardt's 142? We wish the authors had explored other methods of making cross-circuit comparisons, such as normalizing the judges' productivity scores by taking into account differences across circuits in both the means and distributions of published opinions per judge.

44. Choi & Gulati, *supra* note 1, at 45.

B. Opinion Quality

Choi and Gulati try to capture something about the quality of judicial opinions by counting and comparing the number of times a judge's opinions are cited. Of course, not all citations are equal—citations to an opinion may be positive or negative, and may be for a trivial point or a weighty one. In addition, variations across circuits in the number and type of cases heard and norms about publication might influence citation counts irrespective of the quality of a judge's published opinions. To their credit, the authors acknowledge these differences and present a variety of different ways of analyzing citation counts that allow us to see whether and how the ranks will be affected if, for example, we count all citations, or only positive ones.

Citation counts are not a direct measure of opinion quality and Choi and Gulati fully acknowledge this. But they argue that opinions that explain the law better are more likely to be cited, making citation counts a reasonable "market measure" of opinion quality.[45] What they fail to take account of, however, is that judges may cite other judges for strategic or ideological reasons, rather than the persuasiveness of their opinions. On the first point, ambitious judges may cite certain prominent judges strategically—for example, as a signal that they, too, are a certain type of judge. As to the second, suppose that conservative judges are more likely to cite conservative judges, and liberal judges to cite liberal judges, simply because they tend to agree with their substantive outcomes. If, as many argue is the case, the federal judiciary is becoming increasingly conservative, then conservative judges may have higher citation rates simply because they have more like-minded colleagues on the bench than their liberal counterparts. If this is the case, citation counts may be a systematically biased measure: judges in the ideological majority will always tend to be cited more often, and thus win this round of the tournament.

As it turns out, when we compared citation ranks and political party we found that the published opinions of judges appointed by Republican presidents are cited significantly more often those written by Democratic appointees.[46] Whether this is a result of ideological bias could be explored empirically. For example, the authors could examine whether judges tend

45. *See* Choi & Gulati, *supra* note 1, at Part III.B.

46. For Choi and Gulati's Table D data (citations), $p = 0.048$ for a Mann-Whitney test. *Id.* at 94 app. tbl. D. For their Table E data (outside circuit court citations), $p = 0.041$ for a Mann-Whitney test. *Id.* at 100 app. tbl. E. We coded a Republican appointee as 0, a Democrat as 1.

to cite ideologically similar judges more often than ideologically distant ones. We would have liked to see Choi and Gulati analyze the effect of partisanship before assuming that frequent citations are a measure of opinion quality rather than merely a reflection of ideological agreement.

Even if appellate judges do not currently cite like-minded judges more frequently than those more ideologically distant, we wonder whether citation counts would continue to be a reliable measure of opinion quality if the tournament system proposed by the authors were actually implemented. Knowing that citation counts will be taken as a measure of opinion quality, appellate judges might try to game the system, strategically citing ideologically similar colleagues, or engaging in a judicial form of log-rolling ("I'll cite you if you cite me"), in order to enhance the likelihood that they or like-minded colleagues will have a better chance of being selected for the Supreme Court.

C. Independence

The authors' measure of independence turns on a judge's dissents against a judge appointed by a president of the same party ("same-party dissents" for ease of reference). More specifically, they assert that independence is best measured by the difference between the actual proportion of same-party dissents and the expected proportion of same-party dissents (calculated as the proportion of same-party appointees on that circuit).[47] The closer that difference is to zero, the more "independent" the judge under their measure.[48]

We are not sure that this statistical measure actually captures the notion of independence as they have defined it. Suppose, for example, that a judge's dissents are overwhelmingly against same-party appointees. Such a judge would seem to be "independent" according to their definition (independent "from one's political sponsors"[49]), but would score quite poorly on their independence measure, because the actual proportion of same-party dissents would differ greatly from the expected proportion.

If what the authors are trying to capture is the willingness of judges to go against ideologically like-minded colleagues, then it would seem to matter whether they dissent more or less often than would be predicted. Based on their own definition of "independence," then, judges with very high positive scores should be ranked most independent, while those with

47. *Id.* at 64.

48. *See id.*

49. *Id.* at 61.

very negative scores should be ranked least independent. (As Choi and Gulati write, "[t]he more negative the independence score, the more aligned particular judges are with their party line."[50]) Reinterpreting the scores in this way would completely alter the rankings. None of the judges that the authors rank as the top twenty most independent judges[51] would even be in the top forty-percent under the revised measure. Instead, Judge Stanley Marcus, currently listed dead-last with a score of 0.542[52] would be ranked first in independence because he has the largest positive score of any judge in the pool.

We do not mean to argue that ranking the judges from most positive to most negative independence score is the best or even a better way of measuring independence—only that it is plausible given Choi and Gulati's own definition of independence. However, the fact that the rankings are so sensitive to a change to another plausible measure of independence puts a heavy burden on the authors to justify the particular measure they have chosen.

Even accepting that the proportion of dissents against ideologically similar judges tells us something about independence, we question whether the party of the appointing president is the appropriate measure of ideology. Choi and Gulati implicitly assume that all Democrats are equivalently liberal and all Republicans are equivalently conservative, but this is demonstrably not always the case. "Presidents of the same political party," as Micheal Giles and his colleagues write, "vary in their ideological preferences. Eisenhower is not Reagan."[53] Nor, might we add, is Alfred Goodwin an Edith Jones, even though Republican presidents appointed both.

Given that there are likely to be considerable ideological differences between judges appointed by different presidents of the same political party, a better way to capture independence would be to score the judges along a spectrum—one that might place a Nixon appointee closer to a Carter appointee than to a Reagan appointee.[54] When we separated the

50. *Id.* at 64.

51. *Id.* at 65 tbl. 9.

52. *Id.* at 108 app. tbl. G.

53. Michael 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) ("Indeed, the empirical record demonstrates that the voting propensities of some Democratic and Republican presidents do not differ significantly.").

54. For such an approach, see Giles et al., *supra* note 53. For an article looking to ideology and party politics and the different empirical findings with the two measures, see C.K. Rowland & Bridget Jeffery Todd, *Where You Stand Depends on Who Sits: Platform Promises and Judicial Gatekeeping in the Federal District Courts*, 53 J. POL. 175 (1991) (investigating district court standing opinions using party and ideology measures).

judges in the tournament pool according to the particular president who appointed them, not just the president's party, we found differences between judges in the different groups. In particular, we found that Reagan appointees tended to rank higher on Choi and Gulati's independence measure than other categories of appellate judges. This finding raises questions about their measure of independence because political scientists who systematically assess ideology have found Reagan to be the most "extreme" of any president serving over the last six decades.[55] Did Reagan nevertheless appoint judges who are truly more independent, that is, less ideological, than other appellate judges? Perhaps so, but before we accept this conclusion, we would like to know that taking into account critical intraparty distinctions would not make a material difference in the independence rankings, as we suspect they might.

III. CONCLUSION

We found Choi and Gulati's judicial tournament to be an intriguing and fun idea. In fact, we hope the authors consider running additional tournaments and judicial games on the web—perhaps with enthusiasts and admirers rendering votes for their favored contestants. But we must admit, as a proposal for policy reform, we believe it fraught with peril. We are particularly troubled by the fact that Choi and Gulati offer no theory of what makes a great Supreme Court Justice and simply assume the greatest legal minds can be found on the federal appellate courts all over the country. This assumption, and the narrow pool of nominees to which it leads, assures homogeneity. Of course, we recognize that expanding the pool would lead to a total breakdown of the tournament, given the difficulty that the authors (or anyone else) would have in finding measures that predict greatness in diverse venues. But politics being what they are, Choi and Gulati's reform is probably impossible as a positive matter in any event.

55. Virtually all systematic measures of ideology show this to be the case. *See, e.g.*, Jeffrey A. Segal, Richard J. Timpone & Robert M. Howard, *Buyer Beware? Presidential Success Through Supreme Court Appointments*, 53 POL. RES. Q. 557, 561 (2000) (characterizing Reagan as the most conservative president since the Great Depression). Segal et al.'s data on presidential social and economic ideology are available at http://www.sunysb.edu/polsci/jsegal/data/pressc_main.htm (last updated Sept. 22, 2000). *See also* Keith Poole, *DW-NOMINATE Scores* (providing presidential ideology "Common Space Scores"), *at* http://voteview.uh.edu/dwnomin.htm (last visited Nov. 12, 2004).

APPENDIX

Table 1 summarizes the results of sixteen individual studies rating past Justices. The studies have each been given a number, one through sixteen, as designated at the top of the Table. The key to the studies below gives detailed information about each study, in chronological order based on the year of publication. In the Table, an "X" appears for each time a justice was rated "great" by a study. The final column, "Total N," gives the total number of studies rating each Justice as "great."

TABLE 1. Survey of studies rating past Supreme Court Justices as "great"

Nominee	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total N
J. Marshall	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		15
Holmes		X	X	X	X	X	X	X	X	X	X	X	X	X		X	14
Cardozo		X	X	X		X	X	X	X	X	X	X	X	X	X		13
Brandeis			X	X	X	X	X	X		X	X	X	X	X		X	12
Black						X	X	X	X	X	X	X	X	X	X	X	11
Warren						X	X	X	X	X	X	X	X	X	X	X	11
Story	X	X	X	X	X	X		X	X	X				X			10
Frankfurter						X	X	X			X	X	X	X		X	8
Hughes			X	X	X	X	X	X		X						X	8
Taney			X	X	X	X	X	X		X				X			8
Douglas						X	X				X	X	X	X			6
Miller	X		X	X	X	X	X										6
Bradley	X		X	X	X	X					X	X	X		X	X	5
Harlan I				X		X	X	X				X			X		5
Curtis	X		X	X		X											4
Field	X		X	X		X											4
W. Johnson			X	X	X	X											4
Stone				X			X	X								X	4
Brewer	X			X		X											3
Campbell			X	X		X											3
R. Jackson						X	X									X	3
Moody			X	X		X											3
Rehnquist											X			X	X		3
Harlan II														X		X	2
Taft				X									X				2
Blackmun															X		1
Burger															X		1
Butler				X													1
Davis				X													1
Gray	X																1
Jay							X										1
T. Marshall															X		1
Matthews			X														1
McLean				X													1
O'Connor															X		1
Sutherland				X													1
Waite				X													1
E. White			X														1

Key to the studies:

1. CHARLES EVANS HUGHES, THE SUPREME COURT OF THE UNITED STATES (1928).

2. ROSCOE POUND, THE FORMATIVE ERA OF AMERICAN LAW (1938). This study also lists jurists who did not serve on the Supreme Court: James Kent, John Bannister Gibson, Lemuel Shaw, Thomas Ruffin, Thomas McIntyre Colley, and Charles Doe.

3. Felix Frankfurter, *The Supreme Court in the Mirror of Justices*, 105 U. PA. L. REV. 781 (1957).

4. JOHN P. FRANK, MARBLE PALACE: THE SUPREME COURT IN AMERICAN LIFE (1958).

5. George R. Currie, *A Judicial All-Star Nine*, 1964 WIS. L. REV. 3 (1964).

6. Stuart S. Nagel, *Characteristics of Supreme Court Greatness*, 56 A.B.A. J. 957 (1970).

7. SIDNEY H. ASCH, THE SUPREME COURT AND ITS GREAT JUSTICES (1971).

8. ALBERT P. BLAUSTEIN & ROY M. MERSKY, THE FIRST ONE HUNDRED JUSTICES (1978). Blaustein and Mersky asked sixty-five scholars to "grade" the Justices along a continuum from A to E, where A was "great," B "near great," C "average," D "below average," and E "failure." Included here are those rated A.

9. Bernard Schwartz, *The Judicial Ten: America's Greatest Judges*, 1979 S. ILL. U. L.J. 405 (1979). This study also lists jurists who did not serve on the Supreme Court: James Kent, Lemuel Shaw, Arthur T. Vanderbilt, and Roger John Traynor.

10. James E. Hambleton, *The All-time All-star All-era Supreme Court*, 69 A.B.A. J. 463 (1983).

11. Robert C. Bradley, *Who Are the Great Justices and What Criteria Did They Meet?*, *in* GREAT JUSTICES OF THE U.S. SUPREME COURT (William D. Pederson & Norman W. Provizer eds., 1993). Bradley sent surveys to 493 lawyers, judges, scholars, and students. His response rates ranged from thirty-seven percent (scholars) to twelve percent (judges). This column indicates the top choices of all respondents surveyed by Bradley.

12. *Id.* This column indicates the top choices of the scholars surveyed by Bradley.

13. *Id.* This column indicates the top choices of the judges surveyed by Bradley.

14. *Id.* This column indicates the top choices of the attorneys surveyed by Bradley.

15. *Id.* This column indicates the top choices of the students surveyed by Bradley.

16. Comiskey, *supra* note 26. Comiskey surveyed 200 law professors asking them to rank (on a five-point scale ranging from "excellent" to "failure") the fifty-two Justices appointed in the twentieth century (from Holmes to Breyer). Included here are those rated "excellent."