On the perils of drawing inferences about Supreme Court justices from their first few years of service

The ideological boxes into which policy makers, scholars, and lawyers place justices at the time of their appointment are not so tightly sealed.

by Lee Epstein, Kevin Quinn, Andrew D. Martin, and Jeffrey A. Segal

Even before the start of their second year in office, commentators were already reading the tea leaves on Samuel A. Alito, Jr. and John G. Roberts, Jr. According to the prominent legal scholar Erwin Chemerinsky, the two new justices "were every bit as conservative as conservatives had hoped and progressives had feared. [Their] willingness to overrule decades-old precedents certainly gives a sense that major changes are likely ahead in constitutional law in the years to come."[1] Chemerinsky was hardly alone; similar forecasts appeared on the editorial pages of newspapers as ideological disparate as the *Wall Street Journal* and the *New York Times*, as well as hundreds, if not thousands, of blogs across the country.

Forecasting of this sort—a veritable cottage industry each time a new justice completes a term or two of service—may seem harmless enough and sufficiently divorced from the concerns of empirical legal studies to ignore. Nonetheless, the entire enterprise rests on a strong empirical assumption, which, in fact, has important implications for systematic scholarship: that one can draw high-quality inferences about the justices' long-term ideological tendencies from their first few terms in office.

Is this a plausible assumption? Unfortunately, and despite decades of study, we cannot offer a conclusive answer. To some, most notably Hagle, reliance on initial voting records to predict future behavior borders on the absurd.[2] Most justices, he empirically demonstrated, manifest unstable behavior in their "freshman" year relative to the balance of their career. To other scholars, most recently Shipan, predictions based on the first term are not particularly troubling.[3] The instability identified by Hagle, they say, appears insufficiently widespread to be "considered a general phenomenon."[4] In between comes work by Wood et al., which found that roughly half the justices under analysis experienced "acclimation" effects.[5]

In what follows, we hope to bring a fresh eye to this seemingly age-old but nonetheless on-going debate.[6] Drawing on methodological strategies that we used to examine ideological drift on the Supreme Court,[7] we first contemplate the literature's primary concern: to what extent do new justices evince different or unstable behavior in their first year relative to all others? Or, to frame it in more contemporary terms, to what extent can we reach high-quality inferences about the justices' long-term ideological preferences based on one-year's—the first-year's—worth of observations?

Finding that Hagle and others in his camp have the better case—all but 4 of the 26 justices we investigated exhibited statistically significant ideological drift from their *initial* preferences—we turn to questions of substantive importance. Specifically, we demonstrate that movement away from first-year ideal points occasionally manifests itself in consequential doctrinal change—so consequential that a vote in favor of, say, restricting privacy rights or permitting prayer in school in the first term might translate into a vote in opposition before the justice concludes his or her first decade of service. These findings, compatible with our previous work documenting

We thank the Center for Empirical Research in the Law, the National Science Foundation, and the Northwestern University School of Law Beatrice Kuhn Research Fund for supporting our research. Quinn also wishes to acknowledge the Center for Advanced Study in the Behavioral Sciences for its hospitality and support. The project's web site [http://epstein.law.northwestern.edu/research/firstyear.html] houses a full replication archive, including the data and documentation necessary to reproduce our results. Please send all correspondence to Andrew Martin, Washington University School of Law, Campus Box 1120, St. Louis MO 63130. admartin@wustl.edu; 314.935.5863.

1. Erwin Chemerinsky, *The Kennedy Court: October Term 2005*, 9 GREEN BAG 335, at 346 (2006). Our emphasis.

2. Timothy Hagle, *Freshman Effects for Supreme Court Justices*, 37 AM. J. POL. SCI. 1142 (1993).

3. Charles R. Shipan, *Acclimation Effects Revisited*, 40 JURIMETRICS J. 243 (2000).

4. *Id.*

5. Sandra L.Wood, et al., *'Acclimation Effects' for Supreme Court Justices: A Cross-Validation, 1988-1940,"* 42 AM. J. POL. SCI 690 (1998).

6. *See, e.g.,* Mark S. Hurwitz and Joseph V. Stefko, *Acclimation and Attitudes: 'Newcomer' Justices and Precedent Conformance on the Supreme Court*, 57 POL. RES. Q. 121 (2004).

7. Lee Epstein, Andrew D.Martin, Kevin Quinn, and Jeffrey Segal, *Ideological Drift on the U.S. Supreme Court*, 101 NW. U. L. REV. 1483 (2007).

extensive ideological movement on the part of Court members throughout their careers,[8] continue to demonstrate that the ideological boxes into which policy makers, scholars, and lawyers place justices at the time of their appointment are not so tightly sealed.

Inferring future behavior

Inferring future behavior from a justice's first few years in office hardly began with John Roberts and Samuel Alito; in fact, the legal historian George L. Haskins supplies compelling evidence that the practice dates nearly as far back as the Court itself.[9] Surely, though, the most (in)famous modern-day example came after Harry A. Blackmun's first year of service. The Minneapolis-born Blackmun was so closely aligned with his boyhood friend, the conservative Chief Justice Warren Burger, that commentators of the day tagged him a "Minnesota Twin."

While this turned out to be a stunning misnomer, it hardly put the brakes on future forecasting. Quite the opposite. Since Blackmun, virtually every justice has been the object of prediction. After examining Lewis F. Powell's first term in office, one columnist opined that the new justice is emerging "as the conservative's strong man," who will ultimately help "write some pretty good law."[10] Five years later, Justice Scalia was quickly branded Chief Justice Rehnquist's "ideological compatriot."[11] Of David Souter, the chair of the Democratic National Committee declared that he "is slowly demonstrating his loyalty to Republican extremism."[12] Analysts too tried to draw inferences about Justice Kennedy from his first-year record, even though Kennedy (much like Alito) had served for only six months. Bruce Fein, the conservative commentator, declared "there was a clear showing that Justice Kennedy will be in the conservative bloc."[13] The editors of the *New York Times* agreed, as did a host of other longtime Court observers.[14]

Nonetheless, it is hard to deny that tasseography may have reached new heights with Roberts' and Alito's ascensions to the high Court. What with bloggers now joining journalists and scholars, forecasting future ideological tendencies based on a year or less of service has become something of a cottage industry. And a fast-moving cottage industry at that. Attracting substantial attention just a year after Alito and Roberts took their seats was Greenburg's *Supreme Conflict*, famous for its punchline that the two Bush appointees have succeeded in moving the Court to the right.[15]

But what inferences can we really draw? Linda Greenhouse, the astute *New York Times* reporter, suggests we should refrain from the practice altogether. "A Justice's first year on the Court, or even first few years," she once observed, "are notoriously poor indicators of that Justice's eventual role"[16]. No doubt, a non-trivial fraction of justices would agree. As William J. Brennan, Jr. once said,

> There is nothing that you do that prepares you for this job. Even Felix Frankfurter used to say that his lifelong study of the Court never really prepared him for this job. You simply cannot study it from afar and expect to know it. You simply cannot know how you will respond to the legal issues as a Justice, as opposed to a law professor, or a judge on a court of appeals or even a state supreme court. *I know that was certainly true of me.*[17]

Social scientists are less certain but hardly for lack of effort. In a line of inquiry dating back to the 1950s,[18] though picking up considerable steam in the 1970s into the 2000s, nearly countless analysts sought to determine whether new justices evince different or unstable behavior during their first year or so—a period of acclimation on the Court—than in the years to follow.[19] Their findings, as we noted at the outset, have been mixed to say the least. Even if we confine ourselves exclusively to studies of voting,[20] empirical results run the gamut from nearly complete agreement with the Greenhouse/Brennan sentiment to nearly complete disagreement.[21]

Why the mixed findings is a question with many possible answers.[22] Far more relevant here is a commonalty, not a point of distinction, among the existing work: Regardless of their conclusions, recent disciplinary developments have rendered features of the extant studies problematic at best and obsolete at worst. Of particular concern are the measurement strategies they deploy; we are also troubled by their approach to research design.

Let us elaborate, beginning with design—specifically with the issue of how to determine whether new jus-

8. *Id.*

9. George Lee Haskins, FOUNDATION OF POWER: JOHN MARSHALL, 1805-15. VOL. II OF HISTORY OF THE SUPREME COURT OF THE UNITED STATES 152. (New York: Macmillan, 1982).

10. James J. Kilpatrick, *High Court: Proof the System Works*, Los Angeles Times, July 23, 1972, at F6.

11. Charles M. Haar and Jerald S. Kayden, *Private Property vs. Public Use*, New York Times, July 29, 1987, at 23A.

12. Quoted in Steve Daley, *Friends, Foes Hunting Thomas' 'Paper Trail'*, Chicago Tribune, July 4, 1991, at 1.

13. Quoted in Joseph Tybor and Glen Elsasser, *Judging Rehnquist's Cour*t, Chicago Tribune, July 3, 1988, at 1.

14. *See, e.g.,* Stuart Taylor, *High Court Rulings Hint Move to Right*, New York Times, July 3, 1988, at 1; Lyle Denniston, *This Term, Supreme Court Had New Look*, St. Petersburg Times, July 3, 1988, at 1A.

15. Jan Crawford Greenburg, SUPREME CONFLICT: THE INSIDE STORY OF THE STRUGGLE FOR CONTROL OF THE UNITED STATES SUPREME COURT (New York: Penguin Press, 2007).

16. Linda Greenhouse, *The Conservative Majority Solidifies*, New York Times, June 30, 1991, at 1, Sec. 4.

17. Quoted in Jeffrey T. Leeds, *A Life on the Court*, New York Times, October 5, 1986 (Sunday magazine). Our emphasis.

18. Elois Snyder, *The Supreme Court as a Small Group*, 36 SOC. FORCES 232 (1958). *See also* J. Woodford Howard, *Justice Murphy: The Freshman Years*, 18 VAND. L. REV. 473 (1968); John D. Sprague, VOTING PATTERNS OF THE UNITED STATES SUPREME COURT: CASES IN FEDERALISM, 1889-1959 (New York: Bobbs-Merrill, 1968).

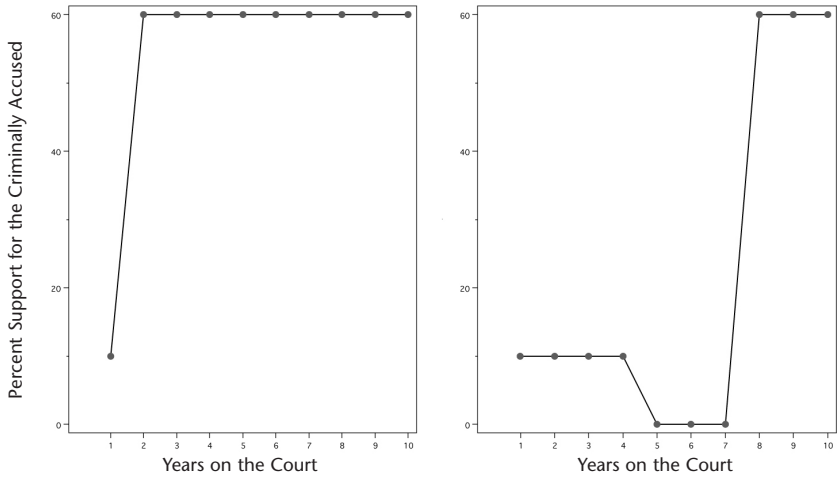
19. *See, e.g.,* Terry Bowen, *Consensual Norms and the Freshman Effect on the United States Supreme Court*, 76 SOC. SCI. Q. 222 (1995); John M. Scheb and Lee W. Ailshie, *Justice Sandra Day O'Connor and the Freshman Effect*, 69 JUDICATURE 9 (1985); Thea F. Rubin and Albert P. Melone, *Justice Antonin Scalia: A First Year Freshman Effect?*, 72 JUDICATURE 98 (1988); Albert P. Melone, *Revisiting the Freshman Effect Hypothesis: The First Two Terms of Anthony Kennedy*, 74 JUDICATURE 6 (1990); Wood et al., *supra* n. 5; Hurwitz and Stefko, *supra* n. 6; Shipan, *supra* n. 3; Hagle, *supra* n.2.

20. To determine if new justices exhibit acclimation or newcomer effects, scholars have looked at a wide range of behavior (e.g., opinion writing, bloc formation, and adherence to stare decisis). While interesting, our emphasis here is on voting behavior, as it is in most contemporary studies

21. For literature reviews, see, e.g., Hagle, *supra* n. 2; Shipan, *supra* n. 3; Hurwitz and Stefko, *supra* n. 6; Christopher F. Smith and S. Thomas Read, *The Performance and Effectiveness of New Appointees to the Rehnquist Court*, 20 OHIO N. U. L. REV. 205 (1993).

22. *See, e.g.*, Hagle, *supra* n. 2; Shipan, *supra* n. 3.

Figure 1: Hypothetical voting patterns of two justices in constitutional criminal procedure cases.



Justice L, depicted in the left panel, cast 10 percent of her votes in favor of defendants in her first year and 60 percent in all others. In his first term, Justice R, in the right panel, also cast 10 percent of his votes in support of defendants. The mean for the rest of his career is 20 percent.

tices are unstable in their voting. While approaches vary from study to study, a prominent one is to compare, say, the percentage of conservative votes cast in the justice's first term, with the mean percentage of conservative votes cast thereafter.[23] If a difference emerges, then analysts claim they have found evidence of a freshman effect—meaning, to use our terminology, that it would be a mistake to draw inferences about the long term based on observations from the first year.

Because this design has its share of intuitive appeal, we understand why scholars have invoked it. Nonetheless, as Shipan correctly observed, it can inadvertently lead to rather severe errors of inference. To see the why, consider Figure 1, in which we present the voting patterns in the area of constitutional criminal procedure for two hypothetical justices, L(eft Panel) and R(ight Panel). Notice that the left panel appears to present few problems for the conventional approach: Because Justice L cast only 10 percent of her votes in favor of defendants in her first term and 60 percent in all others (p .05), scholars would likely conclude that L's first-term behavior reveals little about the balance of her career.[24]

But what of the right panel? Would we be able to make high-quality inferences about Justice R's career based on his freshman preferences? Under conventional methods, the answer is yes: Because no significant difference emerges between R's voting in Term 1 (10 percent) versus the mean of all others (20 percent), first-year behavior provides a plausible predictor of future behavior. The obvious problem with this, the conventional response, is that it fails to attend Justice R's dramatic move to the left even before the end of his first decade of service. Only by developing a design plan capable of capturing the fact that for both R and L evidence of unstable behavior emerges—in fact, evidence sufficiently ample to suggest that inferences based on the first year would provide misinformation about overall career patterns—can we hurdle this obstacle.

Momentarily, we propose such a plan. For now, though, it is worth considering a second, even more fundamental concern with the existing literature—one that Figure 1 also brings to light. Note that in creating the panels we follow the conventional strategy, and depict votes as raw, term-by-term percentages (here, percent support for defendants). That approach, however, fails to attend to changes in the content of litigation ("case stimuli") over time.[25]

Conceptually, the problem is straightforward enough. Absent controls for case stimuli we might incorrectly assume that a justice evinces ideological change after her first term when it well may be the content of the litigation that changed.[26] As a result, we cannot attribute the alteration's source, whether to instability during a justice's early years or to the cases themselves. Greenhouse recognized as much when she wrote of the new justice David Souter,

> While [he] appeared firmly in the conservative camp this term, that impression is due in part to the unusually high proportion of the term's cases that dealt with criminal law and procedure, the area in which he voted most consistently with Chief Justice Rehnquist."[27]

Greenhouse's commentary well captures the flavor of the problem but, because case stimuli can vary in discrete areas of law, it is likely more severe than even she cast it. To provide but one example, consider a justice—call her Justice M(oderate)—who has served on the Court for two terms. Suppose that in her

23. *See, e.g.*, Hagle, *supra* n. 2; Wood, *supra* n. 5.

24. Actually, given that the percentages displayed in Figure 1 fail to control for changes in case content (i.e., changes in issue or case stimuli), this conclusion may not hold. We return to this point shortly.

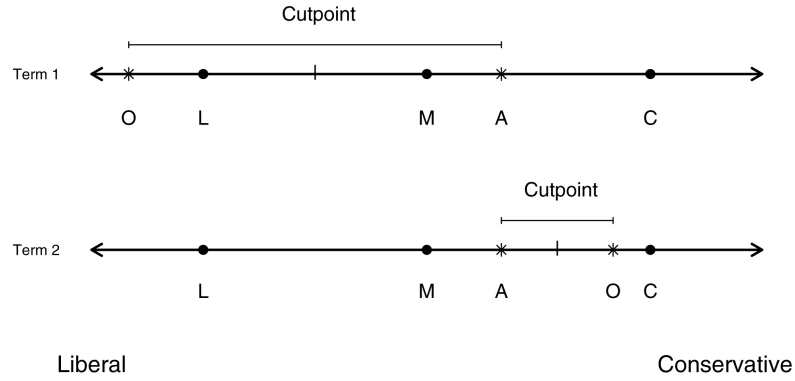
25. We are not the first to recognize this problem. For a review of previous efforts to deal with it, along with their inadequacies, see Shipan, *supra* n. 3.

26. *See, e.g.*, Harold Spaeth, *The Attitudinal Model*, in Lee Epstein (ed.) CONTEMPLATING COURTS. (Washington, D.C.: CQ Press, 1995); Lawrence Baum, *Measuring Policy Change on the United States Supreme Court*, 82 AM. POL. SCI. REV. 905 (1988).

first term, Justice M was quite unsupportive of defendants in Fourth Amendment cases casting only one out of every ten votes in their favor. In the next term, however, Justice M voted to support defendants in nine of ten cases.

If we looked only at these votes, we might conclude that Justice M indeed exhibited unstable, freshman-like, behavior—moving from 10 percent support of defendants in her first term to 90 percent support in her second. But Figure 2 raises another possibility. Here, the horizontal lines represent a single issue, Fourth Amendment search and seizure cases, for two terms (1 and 2). For each term, we have arrayed the ruling of the relevant lower court (labeled O), the alternative policies that could be formulated by a group of justices (labeled A), and the ideal points of hypothetical Justice M, along with Justices L(iberal) and C(onservative).

Figure 2: Hypothetical Fourth Amendment search and seizure cases and justices in ideological space at two time periods (Term 1 and Term 2)



L, M, and C are the ideal points of the three justices—-a liberal, moderate, and a conservative. O represents the status quo policy from a lower court ruling. A is the alternative policy (that could be) formulated by a group of justices. The cutpoint is the location halfway between O and A such that the justices to the left would choose the more liberal outcome, and those on the right the more conservative option.

Suppose that in Term 1 the justices review a lower court decision excluding evidence from a house search on the ground that the judge lacked probable cause to issue the warrant (indicated in Figure 2 by the fairly leftward-located value of O). Further assume, as is standard in the social science literature, that each justice will cast a vote for the more ideologically proximate outcome. If this is so, then—given the choice between the status quo policy O and an alternative policy A, which would, say, allow the introduction of the evidence on a good faith exception—Justice L will vote to uphold the lower court ruling O. Justices M and C, in contrast, will vote for the good faith exception A. Indeed, any justice to the left of the cutpoint—the middle location between the status quo O and alternative A—will vote to uphold, and any justice to the right will vote to overturn.

Now suppose that in Term 2 the lower court rules that a warrantless search is valid, thereby producing the more rightward-located value of O in Term 2. If, once again, the justices make a choice between the lower court ruling O and an alternative A, Justice C will again vote for the more conservative alternative and Justice L for the more liberal one, A. Note, though, that Justice M will now also support the liberal policy, even though her underlying preferences remained stable between the two terms. This implies that just comparing percentages of votes can be profoundly misleading.

A Bayesian dynamic ideal point approach

Until quite recently, overcoming the twin obstacles of design and especially measurement posed a serious challenge to generations of empirical legal specialists.[28] Martin and Quinn have now devised a solution at least to the latter and, we believe, trickier problem of measurement. Using data derived from the votes cast by the justices and a Bayesian modeling strategy, they have generated term-by-term ideal point estimates for all the justices appointed since the 1937 term[29]—estimates that *attend to variation in case content.*[30]

Because the Martin-Quinn (M-Q) Bayesian dynamic ideal point model has been described[31] and applied elsewhere,[32] only a few words are in order here. First, not only does their model allow for ideal points to change over time but the estimates are also directly comparable over time. That is, owing to M-Q's approach, we can compare, e.g., Jus-

27. Greenhouse, *supra* n. 16.

28. Some scholars, most notably Segal, have developed area-specific solutions; in Segal's case, Fourth Amendment search and seizure litigation. See Jeffrey Segal, *Measuring Change on the Supreme Court: Examining Alternative Models*, 29 AM. J. POL. SCI. 461 (1985). But we know of no work that satisfactorily tackles the problem across the range of legal areas.

29. The Martin-Quinn scores, though theoretically unbounded, range from about -6 (Douglas) on the left to about +4 (Thomas) on the right. Across all justices in all terms, the standard deviation of the scores is approximately 2.

30. Andrew D. Martin and Kevin Quinn, *Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999*, 10 POL. ANALYSIS 134 (2002).

31. *See, e.g.*, Martin and Quinn, *id.*; Andrew D. Martin, Kevin Quinn and Lee Epstein, *The Median Justice on the U.S. Supreme Court*, N. CAR. L. REV. 1275 (2005); Epstein et al., *supra* n. 7.

32. *See, e.g.*, Theodore W. Ruger, *Justice Harry Blackmun and the Phenomenon of Judicial Preference Change*, 70 MO. L. REV. 1209 (2005); Barry Friedman and Anna L. Harvey, *Electing the Supreme Court*, 78 IND. L. J. 123 (2005); Paul Wahlbeck, *The Chief Justice and the Institutional Judiciary: Strategy and Constraints on Supreme Court Opinion Assignment*, 154 U. PA. L. REV. 1729 (2006).

tice Souter's revealed preferences in his first term with his second, third, fourth, and so on. Second, as we suggest above, the M-Q method typically will not conflate changes in case content and in ideal points: their model estimates them separately. In other words, using M-Q's approach we can squarely confront the question we pose here—to what extent can we make high-quality inferences about justices based on their first-year record—without having to consider whether any changes we observe are the result of fluctuations in the case content or in the justice's revealed preferences.

Martin and Quinn have dealt with the problem of measurement, but what of design? Certainly, for the reasons we suggest above, we must avoid the extant literature's trap of relying exclusively on a comparison of a justice's revealed preferences (here, measured by the M-Q estimates) in Term 1 with the mean of all others. We rather require an approach that is capable of juxtaposing a justice's initial preferences against the preferences she expresses in each term remaining in her career. Only in this way can we assess the quality of inferences we can draw about a justice's future preferences based on the first years of service.

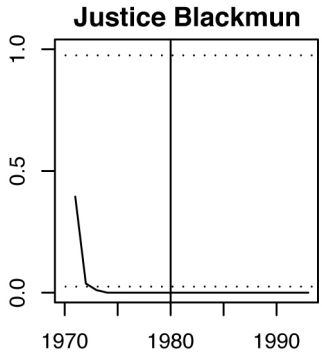
Our solution to this design quandary follows from our work on ideological drift.[33] For each of the 26 justices appointed since 1937 and who served 10 or more terms, we calculate the (posterior) probability that the justice's revealed preferences (i.e., the M-Q ideal point estimates) were more conservative[34] in each subsequent term than her first term.[35] A finding that a justice

33. Epstein, et al. *supra* n. 7.

34. Because the ideal points are continuous, the probability that any two ideal points will be exactly equal is 0. Thus, 1 minus the probability that a justice's revealed preferences are more conservative at some later time is exactly equal to the probability that the justice's later preferences are more liberal. Put another way, knowing the probability of a move to the right allows one to easily calculate the probability of a move to the left.

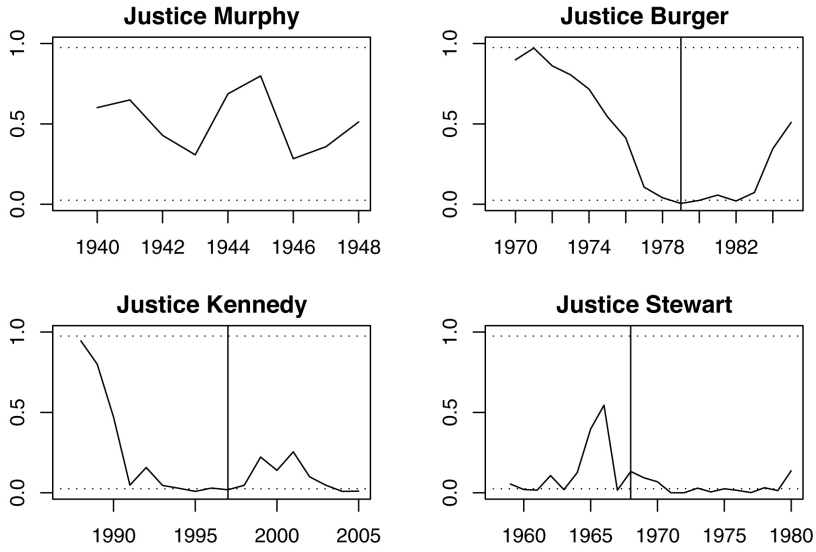
35. In other words, we use the first term as our primary baseline. For more on question of the appropriate baseline, see Hagle, *supra* n. 2.

Figure 3: The probability that Justice Blackmun was more conservative in subsequent terms than in his first term.



The vertical axis denotes the estimated probability. If the solid line is above the top dotted line, then Blackmun was significantly more conservative. If that line is below the bottom dotted line, then Blackmun was significantly more liberal. The vertical line represents Blackmun's tenth year of service.

Figure 4: Four justices, appointed since 1937, who were no more liberal or conservative in more than half their subsequent terms than in their first term.



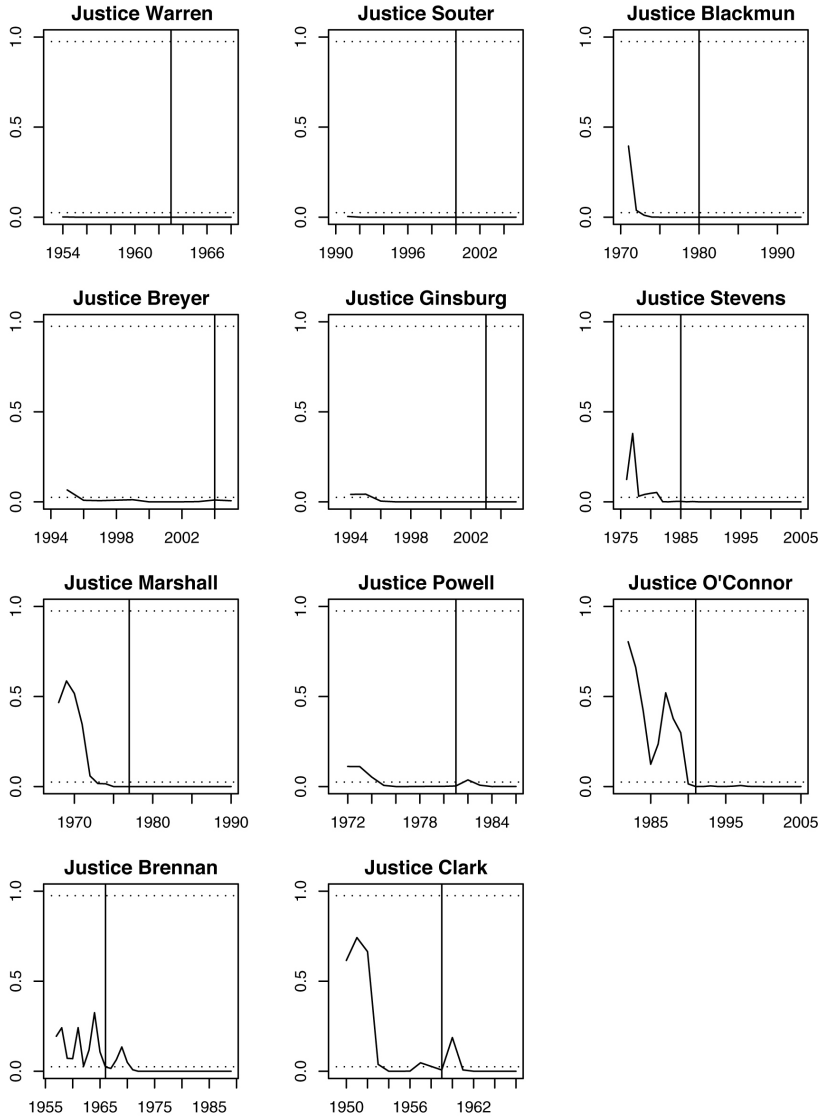
The vertical axis denotes the estimated probability of being more conservative in a future term. If the solid line in each panel is above the top dotted line, then the justice is significantly more conservative. If that line is below the bottom dotted line, then the justice is significantly more liberal. The vertical line within each panel represents the tenth term of service.

exhibits statistically significant differences between her freshman year and more than half of her remaining years ought, we believe, raise serious concerns about the ability to forecast future preferences from the initial term of service.[35]

Applying these calculations to each justice readily enables us to accomplish our primary goal of determining the extent to which she or he, in any given term, exhibited significantly different preferences than in the first year. To supply a straightforward example, consider Figure 3. There we visually depict the results of our calculations for Harry Blackmun, specifically, the estimated probability that he was more conservative (or liberal) in each subsequent term than in his first. Note that if the probability is greater than 0.975 (i.e., Blackmun's ideal point estimate is above the top dotted horizontal line), then we can conclude that he was significantly more conservative in that term than in his first year. Alternatively, if the estimated probability is less than 0.025 (i.e., Blackmun's ideal point estimate is below the bottom dotted line), then we can conclude that he was significantly more liberal in that term than in his first. For purposes of making statistical and substantive inferences, we have added a vertical line representing Blackmun's tenth year of service.

The takeaway from Figure 3 is inescapable: Blackmun moved so far from his first-year ideal point that, despite their best efforts, even the most astute observers of the day could not possibly have drawn accurate inferences about his subsequent behavior based on his first year. Actually, because in no term after his third was he as conservative as he was in his first, even conventional approaches to assessing unstable voting would have picked up the effect; that is, a statistically significant difference emerges between Blackmun's M-Q ideal point estimate in his first term (1.86) and the mean M-Q estimate for the rest of his career (-.19).

Figure 5: Eleven justices, appointed since 1937, who were more liberal in subsequent terms than in their first term.

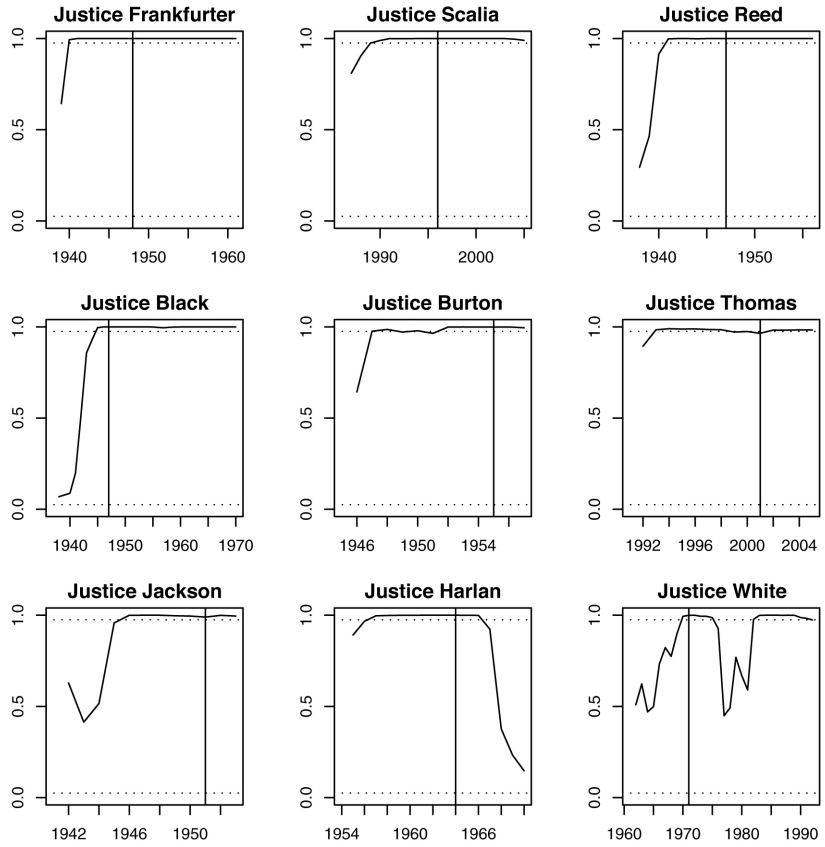


The vertical axis denotes the estimated probability of being more conservative in subsequent terms. If the solid line in each panel is above the top dotted line, then the justice is significantly more conservative. If that line is below the bottom dotted line, then the justice is significantly more liberal. The vertical line within each panel represents the tenth term of service.

36. As we describe in previous work, generating these posterior probabilities presents no major difficulties. It is simply a matter of calculating the fraction of samples generated by the M-Q procedures for which the justice's ideal point in later terms was to the right of the baseline value, the first term. See Epstein, et al., *supra* note 7; Andrew D. Martin and Kevin Quinn, *Assessing Preference Change on the U.S. Supreme Court*, 23 J. L. Econ. & Org. 365 (2007).

37. *See, e.g.,* Ruger, *supra* n. 32; Linda Greenhouse, Becoming Justice Blackmun: Harry Blackmun's Supreme Court Journey (New York: Holt, 2005).

Figure 6: Nine justices, appointed since 1937, who were more conservative in subsequent terms than in their first term.



The vertical axis denotes the estimated probability of being more conservative in subsequent terms. If the solid line in each panel is above the top dotted line, then the justice is significantly more conservative. If that line is below the bottom dotted line, then the justice is significantly more liberal. The vertical line within each panel represents the tenth term of service.

Statistical dangers of drawing inferences

Given the spate of commentary on Blackmun's judicial "journey,"[37] Figure 3 is likely to come to the surprise of no one. But what of the other justices? Is Blackmun the anomaly—the rare justice for whom inferences based on first-year behavior would have been seriously flawed—or the rule? From a statistical vantage point, the answer is clear. By their tenth year of service, 22 of the 26 justices moved, significantly so, away from their first-term ideal point estimate in the majority of their subsequent terms on the Court. Blackmun, in short, is clearly unexceptional.

Let us begin, though, with the exceptions, the four justices—Potter Stewart, Anthony Kennedy, Warren Burger, and especially Frank Murphy—whose observed behavior might encourage the forecasters. While it is true, as we can see in Figure 4, that Stewart and Kennedy eventually grew significantly more liberal, our approach shows that inferences about their ideology based on their freshman term would not have been too far off the mark—at least not for their first decade in office.

The same holds for both Burger and Murphy but their preferences were even more stable.[38] Only in the Chief's ninth and tenth terms—falling near the end of the Carter presidency—was he significantly more liberal than in his freshman year. By the time Ronald Reagan took office, Burger had drifted back to the right, revealing preferences that were statistically indistinguishable from his earlier years. As for Murphy, he represents the true anomaly in our database: the only justice who exhibits no significant preference change between his first year and any of his remaining terms.

In direct juxtaposition come the 22 justices, who—*compared to their first term*—moved to the left, to the right, or in both directions in more than half their subsequent terms. Beginning with the largest subset (see Figure 5), 11 justices, or roughly half those in our database, were significantly more liberal before or at their tenth term than in their first. For the majority, movement came almost immediately. By their second terms, David Souter and Earl Warren, for example, had turned sharply to the left, never to move back. Inferring much about Tom Clark's and, of course, Harry Blackmun's eventual preferences based on their first term too would be deeply problematic in light of the patterns revealed in Figure 5. On the other hand, Justice O'Connor's preferences remained relatively stable until her first decade of service approached, at which point she grew significantly more liberal.

Composing a somewhat smaller, though nonetheless notable group are the nine justices who exhibited

38. Murphy and Burger are the only two justices, who, even under the conventional test, show no signs of voting instability. That is, a comparison of their first-year ideal point estimate and the mean of their remaining terms reveals no statistically significant difference. We cannot say the same for any other justice in our database, including Stewart and Kennedy.

movement in the opposite direction. As we depict in Figure 6, Black, Burton, Frankfurter, Harlan, Jackson, Reed, and Scalia, Thomas, and White were significantly more conservative by their tenth term than in their first. But the trends, once again, differ. For the majority, rightward movement was not long in coming. Frankfurter, for example, was on the Court but two short terms before his revealed preferences grew quite distinct from those in his freshman year. For others, especially White, change came later—closer to the decade mark—but it nonetheless came. Either way, significant movement renders perilous efforts to characterize even their early career stages based on their first years.

By virtually all accounts, the two remaining justices, William O. Douglas and William H. Rehnquist, represented polar extremes on the Court and, yet, they evince remarkably similar patterns. Relative to their first year, as Figure 7 indicates, both were more conservative in at least some of the years prior to their tenth. And, by their eighteenth, both were significantly more liberal relative to their freshman year. In other words, for only a few terms apiece were their revealed preferences indistinct from their first-year behavior.

Substantive dangers of drawing inferences

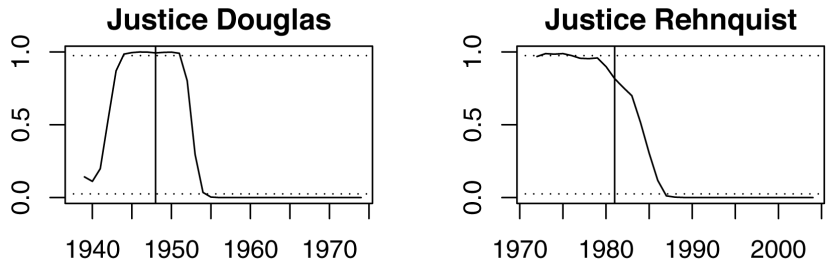
From a statistical vantage point, it would be difficult to observe the patterns in Figures 4 through 7 and reach any conclusion other than the one we have stressed throughout: Regardless of whether we analyze preferences on term-by-term basis, as we have done in the figures, or via the more conventional approach,[39] most justices reveal behavior in subsequent years that differs significantly from their first.

But are the statistical findings compelling enough to dissuade commentators from reaching inferences about Alito, Roberts, and any future justices? Likewise, are they sufficiently persuasive to dispel any doubts about unstable behavior among the justices?

Because statistical significance may not translate into substantive importance, we suspect the answer in both instances is no. Scalia provides a case in point. To be sure, as Figure 6 indicates, he was significantly less

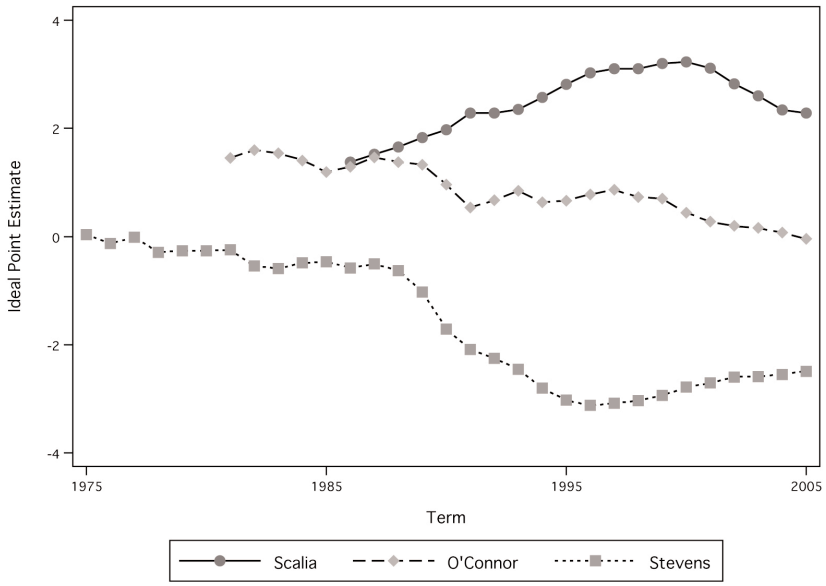
39. *See id.*

Figure 7: Two justices, appointed since 1937, who were both more conservative and more liberal in subsequent terms than in their first term.



The vertical axis denotes the estimated probability of being more conservative in subsequent terms. If the solid line in each panel is above the top dotted line, then the justice is significantly more conservative. If that line is below the bottom dotted line, then the justice is significantly more liberal. The vertical line within each panel represents the tenth term of service.

Figure 8: Justices Scalia's, O'Connor's, and Stevens' estimated ideal points.



The vertical axis is the justice's estimated ideal point. Higher values are more conservative.

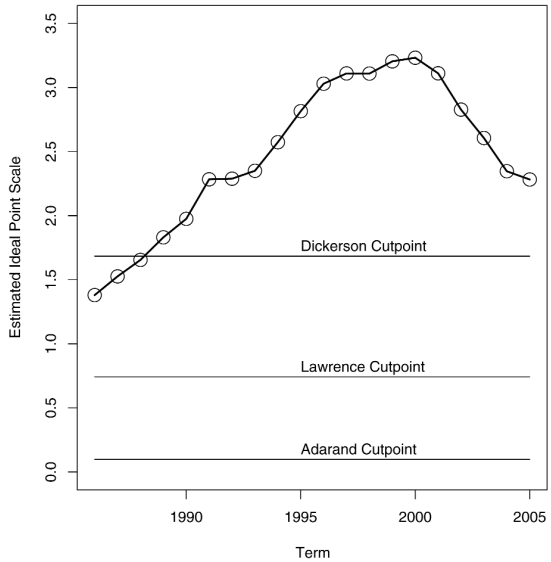
conservative in his first term than in the balance of his career. But we cannot say that commentators of the day were in error to tag the justice a "conservative," and Figure 8 shows why. There we display Scalia's term-by-term (Martin & Quinn) ideal point estimates. (For purposes of comparison, we also plot the estimates for O'Connor, a more moderate justice, and for Stevens, a more liberal justice.) To be sure, Scalia drifts to the right, but only in his first two most "liberal" terms does he even approximate O'Connor (at her most conservative); and his liberalism comes nowhere near Stevens', even in Stevens' early, most right-leaning days.

Antonin Scalia, in short, revealed conservative preferences at the outset of his career, and simply grew more reliably conservative with time. Even more to the point, Scalia's rightward departure was likely of little substantive (doctrinal) consequence, as Figure 9 depicts. Here we again plot Scalia's term-by-term ideal point estimates, along with the cut point lines for three cases implicating different areas of the law: *Dickerson v. United States* (holding that *Miranda v. Arizona* was a constitutional decision that Congress could not overrule by simple legislation); *Adarand Constructors v. Pena* (ruling that all racial classifications must be subjected to strict scrutiny); and *Lawrence v. Texas* (striking down sodomy laws). The cut points provide information about the likely behavior of justices above and below it, such that if a justice's ideal point is above the line, the probability is greater than .50 that she or he will cast a conservative vote (i.e., against *Dickerson* and *Lawrence*, and in favor of *Adarand*).[40]

In the case of *Adarand*, we know

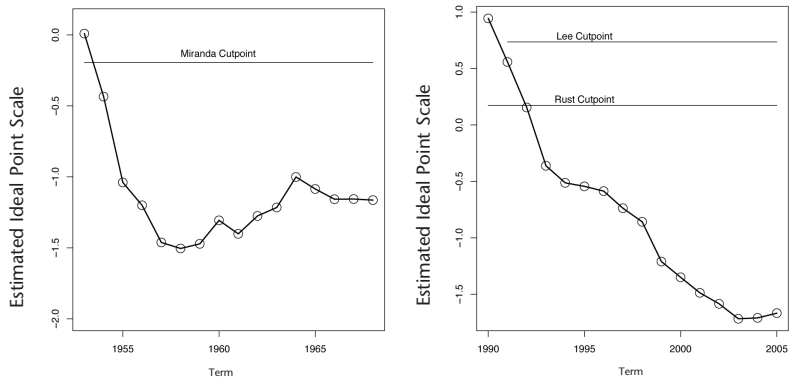
40. We derive these cut points using the Quinn-Martin method. Under their approach, the data and modeling assumptions determine the joint distribution of the ideal points and the cut points. While this joint distribution is large and complex, it is possible to use the conditional distributions of the ideal points—given the cut points—and the cut points—given the ideal points—to fit the model. For more details, see Quinn and Martin, *supra* n. 36; Epstein, et al., *supra* n. 7.

Figure 9: Time series plot of Justice Scalia's estimated ideal points, 1986-2005 terms.



The vertical axis is the estimated ideal point scale, such that higher values are more conservative. The horizontal lines are the cut points for *Dickerson v. United States, Adarand Constructors v. Pena*, and *Lawrence v. Texas* such that points above the line indicate a probability of greater than .50 of voting conservatively; those below the line indicate a greater than .50 probability of voting in the liberal direction (as the Court did in *Dickerson* and *Lawrence* but not in *Adarand*).

Figure 10: Time series plots of Justice Warren's (left panel) and Justice Souter's (right panel) estimated ideal points.



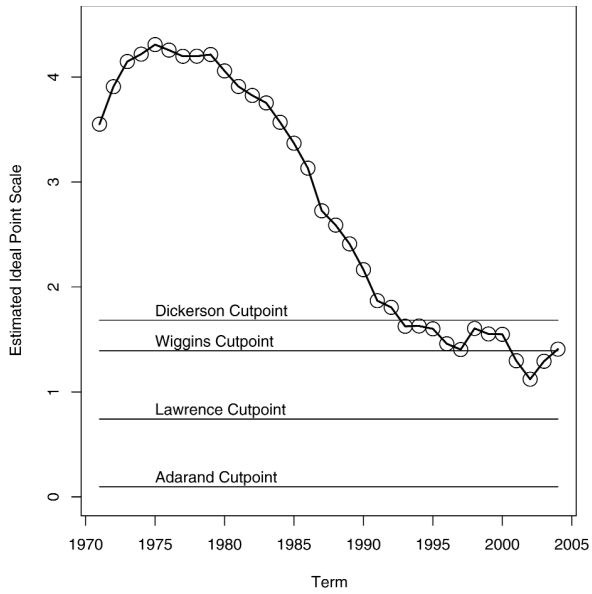
The horizontal lines are the cut points for *Rust v. Sullivan, Lee v. Weisman*, and *Miranda v. Arizona* such that points above the line indicate a probability of greater than .50 of voting for the government (as the Court did in *Rust*); those below the line indicate a greater than .50 probability of voting for the defendant/plaintiff (as the Court did in *Lee* and *Miranda*)

that five justices ruled against the government's set-aside program; we also know that Scalia was among this group. His estimated ideal point in 1995 was above the line and, in fact, he was in the majority in *Adarand*. But also note the location of his ideal points in all previous terms, including in his freshman and sophomore years. Because they are above the cut point, we can safely conclude that even at his most moderate moment—coinciding with the start of his tenure—Scalia would likely have voted to apply strict scrutiny to all racial classifications formulated by the government. More generally, in looking at all three cases depicted in Figure 9, in only *Dickerson*—and for only three terms at that—would we predict a different response had the case come earlier in Scalia's tenure.

Of course we have not scrutinized the cut points of all 1,937 cases resolved since the 1986 term when Scalia joined the Court.[41] But we suspect that additional analyses would only confirm the basic lesson of Figure 8. Because Scalia was sufficiently conservative in his preferences from the start of his service, his turn to the right corresponds to only a marginal change in his jurisprudence. For Scalia, to put it somewhat differently, analyses of statistical and substantive significance depart. The latter implies that any inferences based on Scalia's first term are not necessarily as defective as the former suggests.[42]

For the balance of our justices, however, statistical significant movement, in all likelihood, led to important doctrinal alterations. To provide a few extreme examples, consider Figure 10. There we display the ideal points for Souter and Warren—two justices who made 180-degree turns from the preferences revealed in their first few terms. We also show the cut point lines for three decisions, all in the areas of rights and liberties: for Souter, *Rust v. Sullivan* (upholding regulations that prohibit the use of public funds for abortion counseling) and *Lee v. Weisman* (prohibiting a clergy-led prayer during a public high school graduation); for Warren, the landmark *Miranda v. Arizona*.

Figure 11: Time series plot of Chief Justice Rehnquist's estimated ideal points, 1970-2004 terms.



The horizontal lines are the cutpoints for *Adarand Constructors v. Pena, Lawrence v. Texas, Dickerson v. United States* and *Wiggins v. Smith* such that points above the line indicate a probability of greater than .50 of voting conservatively; those below the line indicate a greater than .50 probability of voting in the liberal direction (as the Court did in all but *Adarand*)

Turning first to Souter's panel, note that when the Court decided to uphold the regulations at issue in *Rust*, Souter was in the majority. Given that his revealed preferences for the 1990 term—his first on the Court—were north of the *Rust* cutpoint line, his vote was not a surprise. Nor, for that matter, was his concurrence in *Lee* supporting the Court's (5-4) decision to prohibit clergy-led prayers at public school graduations. His ideal point in the 1991 term was slightly below the *Lee* cutpoint line.

Far more relevant, and troubling for inference, however, is that Souter's dramatic doctrinal move to the left would have led to mispredictions in the case of *Rust* within three terms of the justice's first year and in the case of *Lee*, just one term thereafter. Indeed, if the Court had heard *Lee* during Souter's freshman—instead of sophomore—year, in all likelihood he would have voted to uphold the prayer. That vote, in turn, may have been enough to convert a five-person majority to strike the prayer into a five-person majority to uphold it (a step many commentators of the day had expected the Court to take). With little doubt forecasts of Souter's "loyalty to Republican extremism" based on his initial period of service were doomed to statistical and jurisprudential fail-

41. We derive the figure of 1,937 from the July 2007 release of Harold J. Spaeth's U.S. Supreme Court Judicial Database, with analu=0 and dec type=1, 6, or 7.

42. We could probably say the same of other extremists who grew only more extreme over time (e.g., Brennan and Marshall). For these justices, as for Scalia, their ideological transformations, from liberals (conservatives) in their first year to more extreme liberals (conservatives) in later terms, likely failed to translate into consequential doctrinal change.

ure.[43]

The same likely holds about inferences of Earl Warren's preferences in the area of criminal law—or, at minimum, in the landmark opinion, his opinion, in *Miranda*. As the right panel of Figure 10 shows, had *Miranda* arrived in his first year, the Chief may well still have written the opinion of the Court, but not one in Ernesto Miranda's favor.

Souter and Warren provide classic examples of the convergence of statistical and substantive significance. Our examination shows that well before their first decade of service, both departed—to a statistically significant degree—from their first year preferences; and our doctrinal analysis in Figure 10 reveals that those statistical departures manifest in changes of consequence—at the least, for several high-profile cases and perhaps for many others as well.

These results are hardly startling: for both Souter and Warren, movement from their first-year ideal points was eventually so dramatic that contemporary scholars hardly missed it. But instability need not be as extreme for substantive change to result. In a paper on the Court's median justice, Martin, Quinn, and Epstein demonstrate as much about Sandra Day O'Connor's gradual drift to the left (also documented in Figure 5).[44] They show that had Justice O'Connor's freshman preferences remained stable, odds are that she would not have provided the fifth vote to uphold Michigan Law School's affirmative action program in the 2003 case, *Grutter v. Bollinger*.

An even more surprising example may be Chief Justice Rehnquist, whose ideal points we depict in Figure 11. Displayed as well are the cut points for *Lawrence, Adarand*, and *Dickerson*, along with *Wiggins v. Smith* (holding that the defendant's attorney had failed to provide effective counsel during the sentencing phase of his capital case). Observe that in neither *Adarand* nor *Lawrence* did Rehnquist's leftward trend translate into doctrinal change: Odds are that at no point in his career would he have voted to invalidate the sodomy law at issue in *Lawrence* or uphold the affirmative action program in *Adarand*. And, in fact, he dissented in both. The criminal cases present a different picture. Had either appeared in Rehnquist's first term, we predict that the then-associate justice would have ruled for the government. It was only in the latter part of his career, when he moved sufficiently to the center, that the odds shifted in favor of the defendant. And, in fact, in both *Dickerson* and *Wiggins* Rehnquist cast votes against the government.

Concluding thoughts

In the conclusion of his important 1993 study, Hagle reported evidence of "significant voting instability" among the justices he examined.[45] While his results have not lacked for challengers, our study provides strong corroborating evidence. Exploiting statistical tools unavailable to Hagle in 1993 or even his adversaries, we find that virtually all Supreme Court justices exhibit significant drift from their first-term preferences; and that drift occasionally manifests in doctrinal change of consequence.

These are the primary lessons of our study, and ones, we hope, that work to resolve a long standing debate in the field. But important challenges remain. While we can say with a high degree of certainty that most justices will move from their first-term preferences, we can specify with no degree of certainty in what direction they will move, nor whether their movement will carry important doctrinal consequences.

Think about the two justices with whom we started our paper, Alito and Roberts. To be sure, our results here indicate that unless either is highly anomalous—a Potter Stewart or Frank Murphy—drawing inferences about their future behavior, as commentators are already doing, is an enterprise doomed to failure. What we do not know is whether the two George W. Bush appointees will come to resemble an Antonin Scalia, that is, a conservative who simply grew more conservative. Or, whether they will follow the path of their predecessors, O'Connor and Rehnquist, and gradually move in the opposite direction from their first-year ideal points[46]—in which case the potential for consequential doctrinal change looms large. Of course, it would loom even larger were the two new justices to morph into Blackmuns, Frankfurters, Souters, Warrens, justices whose first-term behavior actually provided misinformation about their future preferences.

In light of our empirical findings, gaining leverage on these alternatives—whether for Roberts, Alito, or justices past, present, and future—strikes us as a crucial next step. But it is one that requires the development of a theoretical account of why justices move from their initial preferences. Otherwise, speculation on the direction and consequences of instability is just as perilous as inferences that assume a lack of instability.

If we have provided scholars with some incentive to take on the task of developing such an account, as we hope we have, they need not pursue it blindly. Conceptually, several contemporary writers have provided some tantalizing leads. Michael Dorf, for example, argues that justices who come to the Court with executive branch experience are unlikely to move from their initial preferences;[47] and, perhaps not unrelatedly, Lawrence Baum suggests that Republican appointees from outside the beltway are more likely to shift to the left than Republican appointees who are D.C. residents.[48] Another possibility is that justices ascending to the Court from the federal circuits should exhibit relatively weak new-

43. Quoted in Daley, *supra* n. 12.
44. Martin, Quinn and Epstein, *supra* n. 36.
45. Hagle, *supra* n. 2.
46. According to Martin and Quinn's estimates (available at: http://mqscores.wustl.edu/) both Alito and Roberts are to the right of the Court's median, Justice Kennedy.
47. Michael Dorf, *Does Federal Executive Branch Experience Explain Why Some Republican Supreme Court Justices 'Evolve' and Others Don't*, Harv. L. & Pol'y Rev., forthcoming (2007).
48. Lawrence Baum, Judges and Their Audiences: A Perspective on Judicial Behavior (Princeton, NJ: Princeton University Press, 2006).

comer effects. If Baum is right, this effect may be even less still for those coming from the D.C. court of appeals.

Turning to methods, analysts could deploy the data developed here to construct an indicator of the "on average" deviation (from a first-year baseline), thereby adding a "how much" dimension to the "either/or" that we explore in this paper. With this measure in hand, assessing explanations of the influences of variability from freshman behavior should pose few problems.

Whether scholars pursue these leads or develop their own approaches is, of course, less the point than the need to undertake the mission in the first place. Given the empirical findings here, the time seems especially ripe to move beyond description and towards the crucial, yet still unrealized, goal of devising an explanation of instability and its impact on doctrinal development. ⚖

LEE EPSTEIN
is Beatrice Kuhn Professor, School of Law and Department of Political Science, Northwestern University.
(lee-epstein@northwestern.edu)

KEVIN QUINN
is an Associate Professor in the Department of Government at Harvard University.
(kquinn@gov.harvard.edu)

ANDREW D. MARTIN
is a Professor in the School of Law and Department of Political Science, Washington University in St. Louis.
(admartin@wustl.edu)

JEFFREY A. SEGAL
is SUNY Distinguish Professor, Department of Political Science, Stony Brook University.
(jeffrey.segal@stonybrook.edu)