How Judicial Qualification Ratings May Disadvantage Minority and Female Candidates

MAYA SEN, University of Rochester

ABSTRACT

This article uses two newly collected data sets to investigate the reliance by political actors on the external vetting of judicial candidates, in particular vetting conducted by the nation's largest legal organization, the American Bar Association (ABA). Using these data, I show that minority and female nominees are more likely than whites and males to receive lower ratings, even after controlling for education, experience, and partisanship via matching. These discrepancies are important for two reasons. First, as I show, receiving poor ABA ratings is correlated with confirmation failure. Second, I demonstrate that ABA ratings do not actually predict whether judges will be "better" in terms of reversal rates. Taken together, these findings complicate the ABA's influential role in judicial nominations, both in terms of setting up possible barriers against minority and female candidates and also in terms of its actual utility in predicting judicial performance.

I. INTRODUCTION

Despite attempts by presidents and by advocacy groups, federal courts in the United States are still unreflective of the US population. Of the 874 federal judges in service as of 2008, only 24% were women, 10% were African American, and 7% were Hispanic (Just the Beginning Foundation 2012). Fewer than 1% were Asian American, and, even today, there are no federal judges who self-identify as Native American—surprising given the courts' involvement in interpreting federal Indian laws. Among legal actors, politicians, and scholars, there is little dispute that the overall population of female and minority judges falls short of being descriptively representative of the American population at large.

Contact the author at msen@ur.rochester.edu.

I am grateful to Deborah Beim, Matthew Blackwell, Kristin Doughty, Lauren Edelman, David Gelman, Adam Glynn, Jennifer Hochschild, Gary King, Richard Niemi, Kevin Quinn, Susan Welch, Christopher Zorn, and several anonymous referees. Thanks to Catalina Santos, Michelle Pearse, and the Federal Judicial Center for additional research assistance. This research was supported by the Harvard Center for American Political Studies.

Journal of Law and Courts (Spring 2014) 2164-6570/2014/0201-0006$10.00

Compelling explanations of why descriptive representation in the courts has been so difficult to achieve have eluded social scientists, but a possible contributor is thought to be the vetting of presumptive nominees by legal trade organizations such as the American Bar Association (ABA), the nation's largest and most prestigious lawyers' association. For example, according to recent accounts, the ABA preliminarily rejected as "not qualified" 14 of Obama's presumptive judicial nominees. Of these 14 "not qualified" candidates, nine were women, and eight were racial or ethnic minorities; all had their candidacies eventually fail (Savage 2011). The end result, as some commentators have pointed out, is that the ABA now occupies a quasi-governmental role by systematically "vetoing" certain kinds of candidates. Among liberals and minority advocacy groups, the belief is that groups like the ABA are biased against minorities and women. Among conservatives, the belief is that the ABA is biased against conservatives, a notion that has been confirmed by a handful of empirical papers (Lindgren 2001; Lott 2001; Smelcer, Steigerwalt, and Vining 2012).

This article steps into this debate. Looking at newly collected data on the professional and educational backgrounds of the 1,770 individuals nominated to the US district courts since 1960, I find that black and female judicial nominees are indeed more likely to be awarded lower qualification ratings by the ABA, which in turn increases the likelihood that their nominations will fail. I find that this difference persists after matching on education, professional experience, years of legal experience, age, and ideology. The results are also robust to missing data problems associated with confidentially dropped nominations and also to certain kinds of omitted variable bias. Surprisingly, I find no evidence of partisan bias.

To explore the broader implications of this finding, I further examine why minorities and women receiving lower ratings could be problematic. I find that ABA ratings are one of the most predictive factors in whether a judicial nomination is successful, with poorly rated individuals significantly more likely to have their nominations fail. Moreover, in exploring whether these might be a useful predictor of something like judicial "quality" or "performance," I examine a newly collected data set on judges' reversal rates. I find that judges who are poorly rated by the ABA are no more likely to have their opinions overturned than are their higher-rated peers. Taken together, these findings raise questions about why political actors rely on ABA ratings at all. Indeed, the strong reliance on ratings that have little meaningful predictive value of judicial performance suggests that they are used for other reasons. That record numbers of minority and women nominees are currently having judicial candidacies derailed by this vetting process makes this a particularly pressing issue.

This article proceeds as follows. Section II explains how the ABA evaluates nominee qualifications. Section III presents the core hypotheses, while Section IV provides an overview of the data, which are characteristics of some 1,770 individuals formally nominated to the US district courts since 1960. I present the key results showing that women and minorities receive lower ratings in Section V, paying particular attention to

sensitivity to (1) omitted variable bias and (2) selection bias. In Section VI, I turn to exploring why this could present important problems. First, I show that ABA ratings are highly predictive of a nominee's ability to be confirmed, thus raising troubling implications for minorities and women who are more likely to receive lower ratings. Second, I show that ABA ratings are poor predictors of judges' reversal rates, which suggests that their utility in predicting certain aspects of judicial "performance" is quite limited. I conclude in Section VII by discussing the broader implications on judicial selection and on the nominations of minorities and women.

II. HOW THE ABA VETS JUDICIAL CANDIDATES

Once a judicial vacancy arises, the White House—working closely with the Justice Department and (depending on the vacancy) with the senators from the state with the vacancy—develops a list of presumptive nominees drawn from city bar associations, state courts, and area law firms. The short list is then forwarded to the ABA Standing Committee on the Federal Judiciary for additional vetting. No rule exists mandating that presidents must present preliminary lists to the ABA for this "preclearance"; it has nonetheless been a long-standing practice followed since the Eisenhower administration. (A key exception to this has been the George W. Bush administration, discussed below.) Importantly, the list of presumptive nominees is at this point confidential, and the Standing Committee members are prohibited by internal Bar rules from making the names public.

The ABA's Standing Committee then begins reviewing each presumptive candidate's record using three criteria: (1) integrity, which includes "the prospective nominee's character and general reputation in the legal community, as well as the prospective nominee's industry and diligence"; (2) professional competence, which includes "intellectual capacity, judgment, writing and analytical abilities, knowledge of the law, and breadth of professional experience"; and (3) judicial temperament, which includes "the prospective nominee's compassion, decisiveness, open-mindedness, courtesy, patience, freedom from bias, and commitment to equal justice under the law" (ABA 2009).[1] The process by which the ABA assesses these traits is kept confidential, and the committee does not make any ratings public until the president confirms that the presumptive candidate will be put forward as an "official" nominee to the Senate Judiciary Committee (ABA 2009). Thus, many presidents have quietly declined to pursue some number of plausible candidacies, possibly based in part on unfavorable (yet undisclosed) preliminary

1. The committee is composed of 15 individuals from the various federal jurisdictions. This includes the chair of the committee, two members from the large California-based Ninth Circuit, and one member from each of the other 12 circuits. The members are appointed by the ABA president for staggered 3-year terms and cannot serve more than two terms (ABA 2009). Although membership is open to all ABA members, the composition of the Standing Committee has veered toward white and male, with its first African American and female members appointed in 1976 and 1977, respectively.

ABA ratings (Lott 2001). This practice raises the possibility of bias by the ABA when none in fact exists, which is an issue I address below. Once the nominee is rated, the nomination proceeds to the Senate Judiciary Committee, which then can confirm the nominee or effectively kill the nomination (either through inaction or through a vote), in part on the basis of the ABA rating. The nominee may also withdraw.

The opacity of the ABA ratings process has led to assertions that certain candidates are systematically disadvantaged. In this regard, the strongest critique has been that the ABA is biased leftward and that ideologically conservative candidates or candidates nominated by Republican presidents are more likely to receive lower ABA ratings. Examining data from two administrations, for example, Lindgren (2001) finds that confirmed Clinton appeals court appointees with no judicial experience had "9.7 times as high odds of getting the highest ABA rating" as similar George H. W. Bush appointees, controlling for key differences. Although Lindgren (2001) finds no differences between nominees with judicial experience, he does find differences in the criteria that are predictive of high ABA marks under the Clinton and Bush I regimes. (These findings were later critiqued by Saks and Vidmar [2001] on the grounds that the analysis did not include presumptive nominees, as well as district court nominees, and could therefore be biased.) Similar results are obtained by Lott (2001), who collects additional data from presumptive appeals court nominees whose names were not put forward as actual candidates. More recently, Smelcer et al. (2012) use genetic matching to find a bias against Republican appeals court nominees. They find, however, no evidence associated with either race (nonwhite status) or gender.

Comparatively less attention has been paid to the relationship between ABA ratings and race or gender. Lott (2001) notes in passing that African American appeals court nominees, in particular African American Republicans, are most likely to get lower ratings, although these findings do not go to the core of his results (see also Koenig 2012). Similar results are obtained by Haire (2001), who finds that black and female appeals nominees are more likely to get lower ratings, despite controlling for educational and professional differences via multivariate regression. Smelcer et al. (2012), however, find no statistically significant relationship after matching between race or gender and ABA ratings. Anecdotally, however, the belief has increasingly been that the ABA is tilted against some of these candidates, perhaps owing to minorities and women having less trial experience and more government or academic experience (Savage 2011). Informing this belief is additional evidence from sociology and business that minorities and women are evaluated more poorly than their white, male peers in high-level organizations (Bielby and Baron 1986; Fernandez, Castilla, and Moore 2000; Castilla 2008). This is compounded by studies in psychology demonstrating implicit biases against minorities and women by key decision makers in law (Kang 2005; Banks, Eberhardt, and Ross 2006; Greenwald and Krieger 2006; Samuel 2006), public health and medicine (Krieger et al. 2010), business (Bertrand and Mullainathan 2004), and government (Butler and Broockman 2011). Obama administration officials, for example, have been confidentially

informed that the ABA has so far "opposed 14 of the roughly 185 potential nominees the administration asked it to evaluate." Of these, "nine are women—five of whom are white, two black, and two Hispanic. Of the five men, one is white, two are black, and two are Hispanic" (Savage 2011).

This perceived negative treatment of minority candidates has, furthermore, led to tensions between the ABA and Democrats and liberal and racial or ethnic advocacy groups. The Obama administration has declined to pursue the candidacies of some of the presumptive nominees preliminarily deemed by the ABA as being "not qualified," which has led to concerns about the success of its diversity initiatives (Savage 2011). Senator Harry Reid suggested that the ABA "get a new life" after its awarding of a low rating to Obama nominee Gloria Navarro (Tetreault 2010), who was later confirmed by the Senate by a vote of 98–0. And speaking about Latina nominees specifically, one commentator wrote in an opinion piece for the Hispanic National Bar Association:

> I have not seen a single Latina nominee who wasn't either hit or slammed by some establishment group—a bar association, a leader of a not for profit, a bar leader, a judicial committee—as being "intemperate" lacking "seasoning" "inexperienced," "not that bright," etc etc. . . . There's a possibility that the entire cohort of Latina lawyers who want to be federal or state judges just don't deserve it yet, but I'm not buying it. I think there's something else going on, and I think that unearthing what may be going on within the ABA's cloistered process may help us get to the bottom of this. (Raben 2011)

Somewhat less attention has been devoted to the ratings' precise role in the confirmation process and to their ability to predict judicial reversal rates. Within the scholarship on judicial confirmations, Scherer, Bartels, and Steigerwalt (2008) find no predictive power associated with ABA ratings for appeals court judges in terms of the ultimate confirmation decision, although they (and also Allison 1996; Lott 2005) find that low-rated nominees take longer to get through the confirmation process. (Opposite results are presented by Nixon and Goss [2001], who find no delay associated with ABA scores.) Contrariwise, Stratmann and Garner (2004) do find that higher-rated candidates are more likely to be confirmed. In terms of the ABA's ability to predict judicial performance, the evidence is even more scant. Lott (2005) examines the relationship between ABA ratings and appeals judge quality as captured by references in the *Almanac of the Federal Judiciary*, finding no relationship whatsoever between ABA ratings and lawyers' opinions on the judges. De Rohan Barondes (2010) finds, if anything, that more highly rated judges are more likely to be reversed. The only study to find predictive power associated with ABA ratings has been, to my knowledge, Haire (2001), which found that cases written by more highly rated judges were slightly less likely to be reversed. Few of these analyses, however, take into account those selection problems that could substantially skew the results.

III. KEY RESEARCH QUESTIONS

Most studies of ABA ratings have examined the review of candidates at the appeals court level (Lindgren 2001; Lott 2001; Smelcer et al. 2012). I focus here on nominees to the federal district courts, which provides a good basis for understanding whether ratings produced by the ABA and other organizations predict judicial "performance." Of the nearly 300,000 cases per year filed in district courts, around 70,000 are appealed to the US courts of appeals, which then reverse or uphold the district judges' rulings. We can thus determine whether a district court judge's ABA rating will be predictive of his or her reversal rate. In addition, compared to the nine justices serving on the Supreme Court and to the approximately 180 judges serving on the appeals courts, approximately 700 judges serve at any given point on the federal district courts.

Although I examine a different set of judges, the existing literature heavily informs the theoretical expectations. First, we have some limited scholarly evidence that there could be race-based differences in the ABA scores awarded, although the literature is decidedly mixed. Some studies have found a difference (Haire 2001; Koenig 2012); most have not (Lindgren 2001; Smelcer et al. 2012). In all of these studies, however, the focal point was partisan differences rather than race or ethnicity. In addition, recent years have also demonstrated the prevalence of implicit biases, which makes the picture even less clear. This leaves me with a core area of inquiry, which is that white (nonminority) nominees may, on average, receive higher ratings than similarly situated minority nominees. As we have only mixed evidence from the literature, this hypothesis is somewhat tentative and, if it applies, may be the case for African Americans nominees only (Lott 2001). Even more tentative is a related issue concerning male versus female nominees or possible intersectional identities (e.g., those involving minority women). In this regard, there has only been one study finding a difference (Haire 2001), which suggests that possible differences between male and female nominees may exist but that they could be weaker than they are for racial and ethnic minorities. I also note that another expectation is to see some partisan differences between judges appointed by Republicans versus those appointed by Democrats, in line with the existing literature (Lindgren 2001; Koenig 2012; Smelcer et al. 2012).

In addition to these questions, I consider the related issue of why ABA ratings might be important. That is, why would finding systematic discrepancies in the ratings awarded to minorities and women matter at all? I address this question by examining two issues: confirmation and performance. Previous scholarship on the confirmation process agrees that low ABA ratings may have the effect of prolonging confirmation proceedings (Allison 1996; Lott 2005; Scherer et al. 2008), with some studies further finding a decreased probability of confirmation success (Stratmann and Garner 2004). In addition, given the current experiences of Obama nominees and the norm that lowly rated nominees are pulled from consideration, I consider an important possible consequence to be that low-rated nominees will be more likely to have their names withdrawn or their nominations killed by Senate inaction or by Senate vote. That is, being awarded a low

ABA rating will have a negative relationship with the probability that a nominee is confirmed. If minorities are more likely to be awarded these lower ratings, this would then translate directly into a lower probability of confirmation success.

The second point concerns performance. One supposition is that perhaps ABA ratings are invoked by presidents and senators because they believe that ABA ratings gauge accurately, or send some reliable signal about, judicial "quality" and because they think that highly rated nominees are better suited for the judiciary than lower-rated candidates. (Of course, another possibility is that they use ABA ratings for political leverage, which I discuss below.) One such measure of judicial "quality" for district court judges is a judge's reversal rate (i.e., how many of his or her opinions are overturned by higher courts; Choi, Gulati, and Posner 2012; Epstein, Landes, and Posner 2013). This leads to the final hypothesis, for which there is only scant scholarly evidence: ABA scores provide information to political actors about a nominee's quality. Thus, I expect that judges who receive high ABA ratings should have lower rates of reversal than those who receive poor ratings. Finding no relationship between ABA ratings and reversal would raise questions about why they are used at all, especially given the possibility of race- and gender-based discrepancies.

IV. DISTRICT COURT DATA

I investigate these issues by examining (1) 1,652 US district judges confirmed between 1960 and early 2012 and (2) 121 individuals formally nominated to these courts between 1960 and 2012 but whose nominations were ultimately withdrawn, rejected, or killed due to Senate inaction.[2] For the 1,652 district judges actually confirmed, I collected their ABA rating using data from the Federal Judicial Center (FJC). The ABA currently awards three possible ratings: (1) "well qualified," for which "the prospective nominee must be at the top of the legal profession in his or her legal community"; (2) "qualified," in which the nominee "satisfies the Committee's very high standards"; and (3) "not qualified," where "the prospective nominee does not meet the Committee's standards" (ABA 2009). Two other categories have been discontinued: (4) "exceptionally well qualified," discontinued in 1989, and (5) "not qualified by reason of age," discontinued in 1981. Because the "exceptionally well qualified" rating was discontinued, I present many of these analyses collapsing that category into the next-higher category, "well qualified." (The results are unaffected.) Only three confirmed judges ever received the "not qualified by reason of age" rating, which was always awarded to individuals over age 60 at the time of nomination. Because so few nominees received this rating and because this rating was deterministic, I drop these nominees. For the approximately 120 nominees who were

2. I begin with 1960 because the first African American district judge was confirmed in 1961, and there is no support for cross-race comparisons and little support for cross-gender comparisons before 1960.

Table 1. Distribution of ABA Ratings for US District Court Candidates Formally Nominated after 1960

	Not Qualified	Qualified	Well Qualified	Exceptionally Well Qualified	*N*
All	.01	.44	.53	.02	1,776
Whites	.01	.41	.56	.02	1,484
Blacks	.02	.58	.40	.00	163
Hispanics	.02	.59	.38	.01	102
Women	.01	.49	.50	.00	304
Men	.01	.42	.54	.03	1,472
Democrats	.02	.44	.52	.02	788
Republicans	.01	.43	.54	.02	988
Confirmed	.01	.43	.54	.02	1,652
Not confirmed	.07	.54	.39	.00	121

Source.—Federal Judicial Center.

never confirmed, I collected their ABA ratings directly from the ABA or from other sources, such as Goldman (1997).

A demographic breakdown of scores by race, gender, and party affiliation of the appointing president is provided by table 1. Only about 3% were ever awarded the two most extreme categories, "exceptionally well qualified" and "not qualified." About 44% have been awarded the second-lowest category, "qualified," with the majority of judges, 53%, being awarded the second-highest category, "well qualified." The same is, however, not true for minority judges, more of whom were awarded the lower "qualified" category; 58% of African Americans and 59% of Hispanics received this category. The ratings are also on average worse for candidates who were never confirmed, a point to which I return below.

The data from the FJC also include demographic characteristics for confirmed judges, including age, race (with "Hispanic" being a distinct categorization), law school attended, and a brief description of the judge's previous professional experience. Because previous professional experience speaks directly to the ABA's criteria of "professional competence," I used automated content analysis to code these excerpts to indicate whether each nominee (1) was a former law clerk,[3] (2) had ever served as a US attorney or as an assistant US attorney, (3) had worked in the solicitor general's office (as a deputy or assistant solicitor general), (4) had ever served as a state judge (either as a state supreme court or state lower court judge), (5) had ever been a former federal judge (e.g., magistrate, territorial, or bankruptcy judge), (6) had worked as a full-time law professor

3. Coding was automated using the statistical program R. "Former law clerk" was coded as whether the nominee previously clerked for an individual judge, as opposed to serving as a court clerk, clerk of the court, or court staff attorney, occupations that were sometimes designated by the FJC as "law clerk."

or law school dean,[4] (7) had experience as an attorney in private practice,[5] or (8) had ever been a public defender.[6] Finally, because the ABA specifically purports to take into account legal practice and trial experience in its ratings, I also used the automated coding to calculate the years that each nominee had spent at each of these jobs. This allowed me to create a measure of (9) years of legal practice experience, which was the total number of years the candidate spent in private practice, as a US attorney or assistant US attorney, or as a public defender. By no means is this the only way to capture this measure; however, singly including years of experience at each job did not meaningfully alter the results.

I further coded the law school attended by using 2001 *US News & World Report* rankings and dividing them up into (1) elite law schools in the "top 14," (2) other law schools in the top 25, (3) law schools ranked between 26 and 50, (3) law schools ranked between 51 and 76, (4) law schools ranked between 76 and 100, and (5) law schools ranked outside of the top 100. (These are a somewhat rough measure for judges attending law school in the 1960s and '70s; an assuaging factor is that the composition of the top 14 schools has not changed over time.) For those 121 individuals whose nominations were somehow derailed and who are not in the FJC data, I gathered a new data set containing parallel information.[7] For these data, I examined personal biographies, newspaper articles, the Congressional Record, and Department of Justice archives. The breakdown by race, gender, and party affiliation for all nominees is reported in table 2.

V. FACTORS PREDICTIVE OF ABA RATINGS

I begin the empirical inquiry by simply examining which characteristics are predictive of high ABA ratings. The outcome variable here is whether the nominee was highly rated by the ABA, receiving either an "exceptionally well qualified" or "well qualified" rating. (Inferences do not change when using an ordered logit specification.) Thus, table 3 shows the relationship between key professional characteristics and whether a nominee earned one of the higher ABA ratings versus one of the lower ones. I also include dummy variables for race or ethnicity (with non-Hispanic whites as the baseline group), gender, and appointment by a Republican, which are the variables of interest in the analysis. Other controls include the age of nominees at time of nomination, whether they attended a top 14 law school, previous professional experience, and year fixed effects (models 2 and 3). Because the ABA ratings awarded could fluctuate according to intersectional

4. That is, the coding excluded individuals who were identified by the FJC as having worked as "adjunct" professors, "visiting" professors, or other part-time academic designations. Errors in self-reporting to the FJC are presumably random.

5. This was coded as anyone who (1) was designated as at some point having worked in "private practice" or who (2) had listed in the employment text field a private law firm or a private law firm designation in the title (e.g., "partner," "associate").

6. This included people who had the words "federal" or "public defender" in their job title or who worked at a location that contained the words "federal defender" or "public defender" in the title.

7. Due to data availability, this parallel data set does not include years of legal practice experience.

Table 2. Demographics of US District Court Nominees Named after 1960

	All	Whites	Blacks	Hispanics	Women	Men	Democrats	Republicans
Average age at investiture	50.06	50.43	48.55	47.71	47.93	50.50	50.59	49.65
Female	.17	.15	.28	.27	1.00	.00	.24	.11
Nominated by Democrat	.44	.40	.73	.50	.63	.41	1.00	.00
Top 14 law school	.30	.31	.28	.25	.30	.30	.32	.29
Private law school	.52	.51	.66	.44	.59	.51	.55	.51
Law clerk	.21	.22	.14	.11	.35	.18	.23	.20
Law professor	.06	.05	.12	.06	.07	.05	.07	.05
Private practice	.91	.93	.78	.84	.83	.93	.91	.92
US attorney	.08	.09	.03	.05	.06	.09	.06	.10
Assistant US attorney	.20	.19	.28	.20	.29	.18	.19	.21
Justice Department lawyer	.05	.05	.07	.04	.06	.05	.05	.05
Public defender	.05	.03	.12	.15	.07	.04	.07	.03
US magistrate judge	.09	.08	.10	.13	.19	.07	.09	.08
US bankruptcy judge	.01	.01	.04	.00	.03	.01	.02	.01
State judge	.40	.38	.54	.50	.43	.39	.41	.40
Years of practice	15.88	16.76	10.08	12.87	11.02	16.87	15.63	16.07
N	1,789	1,494	163	103	305	1,484	794	995

Source.—Federal Judicial Center.

identities (e.g., being black and female), I also include an interaction between race and gender (models 3 and 4). Limiting the sample to confirmed judges and including the judge's years of legal experience changes little, as does including the judge's judicial common space score, which serves as a measure of political ideology (model 4).[8]

As the results from table 3 suggest, certain traits are positively linked with earning a high ABA score. For example, individuals who have previous judicial experience (e.g., previous experience as a state or federal magistrate judge) are more likely to receive one of the two higher ratings, as are those who have more years of legal practice experience (model 4). Other characteristics that are linked with higher scores include whether the nominee attended a top 14 law school, spent time in private practice, or served as an assistant US attorney. Age, which is measured in years at time of commission or nomination, is also positively associated with receiving a higher ABA score. Two other characteristics—whether the judge was a former law clerk and whether the judge attended

8. The judicial common space score takes advantage of "senatorial courtesy," the long-standing practice of presidents to consult US senators on judicial vacancies in their home states. Thus, the judicial common space scores are based on either the (1) the ideological common space score of the appointing president or, when the senator/s and president are of the same party, (2) the common space score for the senior senator or average of the two senators, if both from the same party (Giles, Hettinger, and Peppers 2001; Epstein et al. 2007).

Table 3. Logit Regression for US District Court Nominations between 1960 and 2002

	Model 1	Model 2	Model 3	Model 4[a]
African American	−.70***	−.89***	−1.11***	−1.50***
	(.19)	(.21)	(.26)	(.33)
Hispanic	−.62***	−.70***	−.92***	−.92**
	(.23)	(.24)	(.33)	(.46)
Female	−.20	−.44***	−.53***	−.88***
	(.15)	(.16)	(.19)	(.27)
Republican	.02			
	(.11)			
Age	.07***	.06***	.06***	.04***
	(.01)	(.01)	(.01)	(.01)
Top 14 law school	.13	.23*	.07	.09
	(.12)	(.13)	(.15)	(.18)
Private law school	.14	.19	.06	−.04
	(.11)	(.12)	(.14)	(.17)
Law clerk	.40***	.19	.19	.25
	(.13)	(.15)	(.16)	(.22)
Law professor	−.11	−.04	−.01	.11
	(.22)	(.24)	(.26)	(.31)
Private practice	.54***	.51**	.53**	.47
	(.19)	(.22)	(.23)	(.35)
US attorney	−.29	−.20	−.23	−.24
	(.19)	(.21)	(.23)	(.28)
Assistant US attorney	.68***	.64***	.65***	.29
	(.15)	(.16)	(.17)	(.21)
Justice Department	.23	.16	−.08	.15
	(.25)	(.28)	(.30)	(.41)
Public defender	.41	.24	.02	.22
	(.26)	(.29)	(.31)	(.47)
Federal magistrate	.95***	.80***	.87***	.52
	(.21)	(.23)	(.25)	(.35)
Federal bankruptcy	.55	.98**	1.13**	1.15*
	(.44)	(.49)	(.54)	(.66)
State judge	.24**	.27**	.17	.36*
	(.11)	(.12)	(.13)	(.18)
African American × female			.18	−1.39
			(.50)	(1.24)
Hispanic × female			.02	.63
			(.57)	(1.04)
Judicial common space score				−.62
				(.54)
Years of practice				.03**
				(.01)
Year dummies		✓	✓	
District dummies			✓	✓
Constant	−4.01***	−17.25	−18.23	−19.72
	(.50)	(535.41)	(1,455.40)	(2,399.54)
N	1,718	1,629	1,629	1,092

Source.—Federal Judicial Center.

Note.—Outcome variable is whether the nominee received a (1) well qualified or exceptionally well qualified ABA rating vs. (2) a not qualified or qualified rating.

[a] Restricted to confirmed nominees only and includes judicial common space score and years of prenomination lawyering experience.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

a top law school—are positively linked with higher ABA ratings but at times fall shy of significance. Thus, we can say that some measures of prestige (e.g., experience as a federal prosecutor) matter, as do certain kinds of legal work experience, including private practice experience and judicial experience.

Three traits are consistently negatively linked with receiving high ABA ratings. These include a nominee being (1) female, (2) African American, or (3) Hispanic. Effects for all three are statistically significant under any model specification, although the interaction of gender and race appears not to be, suggesting no statistically significant differences in how the effect varies between minority men and minority women. (There were insufficient numbers of Asian American or Native American judges to make meaningful inferences about these groups.) A fourth variable of interest, a judge being named by a Republican, appears to have no real relationship with attaining a high ABA rating, and it is never statistically significant under any model specification. (In results not shown, I confirm this nonfinding via matching analyses; under no matching specification are the differences between judges named by Republicans vs. those named by Democrats statistically significant.) A judge's judicial common space score, a more nuanced measure of ideology, is also not predictive of whether he or she receives a high ABA rating.

A. ABA Ratings, Racial Minorities, and Women

The results presented in table 3 suggest that racial/ethnic minorities and women are receiving lower scores, even after controlling for key characteristics. However, because nominees differ in terms of their legal training, professional backgrounds, judicial preparation, and years of experience (table 2), comparisons via multivariate regression may mask substantial differences by making inferences outside of the common support of the data.

1. Matching Analysis

To account for such differences, I rely on matching (Ho et al. 2007). Matching allows the comparison of nominees who are identical across key characteristics. Thus, a female nominee who graduated from a top 14 law school and who previously served as a federal magistrate will be compared with a male nominee who also graduated from a top 14 law school and who also worked as a federal magistrate.

This approach offers several advantages. First, matching is an effective preprocessing step that reduces dependence on statistical modeling assumptions (Ho et al. 2007). Second, and relatedly, matching effectively tests all possible ways that variables could interact with each other. We may think, for example, that the ABA might treat male and female judges differently but only among individuals attending lower-ranked law schools. By pruning the data, matching resolves this problem and isolates the effect of a nominee being female or African American, regardless of the possible ways that other variables may be affecting one another. To implement the matching, I use coarsened exact matching (Iacus, King, and Porro 2009, 2011), which allows exact matching on key variables and

Table 4. Prematching (for All Judges) and Postmatching Characteristics for (1) Blacks Matched to Whites and (2) Women Matched to Men

	Prematch	White	Black	Women	Men
Female	.17	.08	.08	1.00	.00
Appointed by Democrat	.44	.69	.69	.63	.63
Top 14 law school	.30	.28	.28	.46	.46
Law professor	.06	.03	.03	.02	.02
Private practice	.91	.92	.92	.94	.94
Assistant US attorney	.20	.11	.11	.12	.12
Justice Department lawyer	.05	.09	.03	.06	.07
Law clerk	.21	.17	.17	.26	.26
US magistrate judge	.09	.00	.00	.02	.02
US bankruptcy judge	.01	.00	.00	.00	.00
State judge	.40	.64	.64	.43	.43
Average commission year	1988.9	1990.4	1990.1	1994.72	1994.6
N	1,786	78	36	65	141

Source.—Federal Judicial Center.
Note.—Matching was done via coarsened exact matching.

coarsening and then matching approximately on the few variables that are continuous (discussed below). Coarsened exact matching has the advantage of allowing for this approximation to be as close as needed to remove biases. I also have the advantage of matching exactly—the best form of matching—on a large portion of the variables.

Unfortunately, because there are so few Hispanics and nearly no Asian Americans (and no Native Americans), I focus the race/ethnicity part of the matching analysis on African Americans, here compared to whites. In each instance, unless noted otherwise, I first match on the relevant personal and professional characteristics, including (1) nominee gender (or race in the case of women), (2) the identity of the appointing president, (3) age (using four age cohorts), (4) state judge, (5) US attorney, (6) assistant US attorney, (7) assistant or deputy solicitor general, (8) federal magistrate or bankruptcy judge, (9) law professor or dean, (10) private practice experience, (11) public defender experience, (12) law clerk experience, and (13) rank of law school.[9] Next I calculate the difference in means in the two populations (black vs. white nominees, or female vs. male nominees) in terms of the ABA rating awarded. I do so via a simple least squares model; using a more complicated parametric model that controls for the matching covariate results in substantially similar inferences but has the potential to introduce model dependence.

A summary of district court nominee characteristics postmatching is given by table 4. This matched sample of nominees is, as expected, slightly different from the original

9. For this variable, I match according to the rank cohort of law school, as described above in Sec. IV.

Table 5. Change in Probability, after Coarsened Exact Matching, of Receiving One of the Two Highest ABA Ratings

	Change in Probability of High Rating	95% Confidence Interval	*N*
1. Among all nominees:			
African Americans	−.42***	(−.60, −.24)	114
Women	−.19***	(−.33, −.05)	206
2. Among confirmed only (includes ideology and years of legal practice):			
African Americans	−.31**	(−.61, −.02)	44
Women	−.22**	(−.43, −.01)	65

Source.—Federal Judicial Center.

Note.—Row 1 includes all nominees, confirmed and unconfirmed, and matches on professional characteristics and law school attended. Row 2 includes confirmed judges only and additionally matches on judicial common space score and years of prenomination legal experience. For example, African American nominees have a 42 percentage-point drop in the probability of receiving one of the two higher ratings from the ABA, postmatching. All of the postmatching effects are statistically significant at the 5% level.

** $p < .05$.

*** $p < .01$.

prematched sample (the first column of table 4 and as well as table 2) but certainly not fundamentally atypical. Very few of the matched nominees had experience working as magistrate or bankruptcy judges, as law professors, or as assistant US attorneys, a testament to the small number of such individuals in the population of nominees at large. In addition, the matched sample has (in most instances) a greater proportion of individuals who attended a top 14 law school, whose careers were spent in private practice, and who were nominated by Democrats. Finally, the average commission or nomination year fluctuates somewhat from the overall sample, reflective of the fact that certain diverse candidates are more likely to be nominated in later administrations. This further alleviates the potential concern that the previous results were driven exclusively by Carter-era nominees.

Results after matching on these key characteristics are presented in table 5, row 1. Looking at African Americans, an estimate of −0.42 indicates that black nominees are on average 42 percentage points less likely to receive a high rating from the ABA than are professionally identical whites nominated by the same president, a difference that is also statistically significant at the 5% level (with 95% confidence intervals of −0.60 to −0.24). Different coarsening and matching on other professional factors never changes the direction or even rough magnitude of the results. The results attenuate slightly for female nominees. For women, I match them to men across the same characteristics as before; the one exception is that instead of matching on the nominee's gender, I match on the nominee's race or ethnicity so as to hold that constant. The results demonstrate that women are, on average, 19 percentage points less likely than identically situated men to receive one of the two highest ratings from the ABA. These findings are also statistically

significant at the 5% level (with 95% confidence intervals of −0.33 to −0.05). Both sets of results are consistent with the results obtained from parametric methods.

I further consider two additional sources of systematic discrepancies. The first is that African Americans or women may be more likely to be left-leaning, reflected in the identities of the presidents who appointed them, and that this left-leaning tendency could potentially lead to lower ABA ratings. (I note, however, that no study has ever shown a bias against more liberal nominees by the ABA.) The second is years of legal experience. Here, the ABA itself has stated that a nominee "ordinarily should have twelve years experience in the practice of law" (ABA 2009). Thus, I match on whether the judge did or did not have at least 12 years of legal practice experience. These results are presented in table 5, row 2. Although the number of matched observations drops significantly (because these data are available only for nominees who were successfully confirmed), the point estimates waver only slightly. For African Americans, there exists a 31 percentage-point drop in receiving a top score, even when controlling for previous legal experience. For women, we see a 22 percentage-point drop, a larger effect than before. Both of these estimates are significant at the 5% level.

2. *Sensitivity to Omitted Variables*

Although I match on, or otherwise take into account, a substantial number of factors that could possibly influence the ratings awarded, it is possible that (1) we do not have access to the full breath of information available to the ABA's Standing Committee on the Federal Judiciary or that (2) some of the information used by the ABA is inherently qualitative in nature and not included in the amalgam of quantitative data. To gain some traction over the possibility that unobserved covariates are driving the results presented in table 5, I use a method of sensitivity analysis described by Rosenbaum (2002).[10] This sensitivity analysis works roughly by hypothetically "increasing" the level of unobserved characteristics (i.e., omitted variables) in the "treated" population (e.g., racial and ethnic minorities, women) until the results are no longer significant. Thus, the sensitivity analysis gives us an estimate of the size of the omitted variable bias (denoted as Γ) that must be present in order for the results to be called into question. For example, a result of $\Gamma = 1.2$ for African American nominees means that there must be 20% more of some unobserved trait among the African American nominees for the results to lose significance. Although there is no firm agreement in the literature about the minimum Γ value for observational studies, anything above $\Gamma = 1.5$ appears to indicate substantial insensitivity to unobserved variables, while $\Gamma = 1.2$ is around average (Keele 2010). Importantly, sensitivity analyses are limited in that they can only test against sensitivity to omitted variables, and, furthermore, showing that there is a lack of sensitivity is not the same as proving that the previous estimates are correct. However, this type of sen-

10. This is implemented in R using the rbounds package described in Keele (2010).

Table 6. Rosenbaum Sensitivity Analysis Results

	Postmatching Coefficient	p-Value	Γ Statistic	N
1. Among all nominees:				
African Americans	−.04	.00	2.8	114
Women	−.19	.01	2.0	206
2. Among confirmed only (includes ideology and years of legal practice):				
African Americans	−.31	.04	1.2	44
Women	−.22	.04	1.1	65

Source.—Federal Judicial Center.

Note.—Columns display original postmatching estimates, exact p-values assuming no omitted variables, Rosenbaum sensitivity analysis gamma statistics, and number of observations. Row 1 includes results after matching on professional characteristics and law school attended. Row 2 includes confirmed judges only and additionally matches on judicial common space score and years of prenomination legal experience.

sitivity analysis allows us to be skeptical of the modeling assumptions and to check how strong omitted variable bias would have to be to overturn the results.

The sensitivity analyses are presented in table 6 and demonstrate that (by observational standards) the results are robust to unobserved variables. In order to make the results insignificant, some trait would have to be present in the African American nominee population approximately three times as often as in the white population. For women, the results are more sensitive, an outcome consistent with the smaller treatment effect for this group (table 5). In order to make the results insignificant, women nominees would have to have some treatment approximately two times as often as male nominees. Matching on ideology and legal experience reduces the number of observations and, thus, makes the results more sensitive. However, even then, the sensitivity results are in line with what we would expect from most observational studies.

Given that the analysis already controls for clerkship experience, professional experience, quality of legal education, previous judicial experience, ideology, and years of legal practice experience, it seems unlikely—although not out of the realm of possibility—that some set of omitted variables is driving the results. It could be the case that, for example, African American judges are nearly three times less likely than white judges to have been on their school's law review or to have graduated as members of the Order of the Coif, a law school honors society—despite controlling for rank of law school, subsequent clerkship experience, and judicial experience (for which such metrics might be very predictive). In this regard, a lively debate is ongoing about the relative successes of African Americans at the nation's prestigious law schools (Sander 2004; Ho 2005). I do not engage the debate between Sander (2004) and Ho (2005), which turns specifically on causal identification issues, but note that both support the finding that similarly situated African Americans and whites (including those who attend the same law school and have similar grade profiles) tend to pass the bar at similar rates. The data are different, but we would expect that those with similar law school, clerkship, and professional back-

grounds and legal experience would also have similar nominee profiles. Taken together, the evidence therefore counsels away from assuming the results are driven exclusively by omitted variables.

3. *Sensitivity to Selection Bias*

As noted, the ABA makes public its qualification ratings only for those individuals who were eventually nominated by the White House and whose candidacies advanced to the Senate Judiciary Committee; that is, ABA qualification scores are available only for actual nominees, not presumptive nominees (ABA 2009). Thus, any publicly available data systematically exclude ABA ratings of individuals whose candidacies were dropped secretly by presidents during the ABA's confidential "preclearance" stage.

Although not publicized, anecdotal evidence suggests that the number of failed presumptive nominees appears to be around three to five per 4-year term.[11] A significant concern is, however, that not including these individuals in the analysis could bias the results. For example, it could be the case that presidents starting with Jimmy Carter were eager to appoint minority judges, perhaps in order to increase more rapidly the proportion of black and women judges on the courts. Under such a scenario, it could be likely that presidents who had their confidential "short lists" vetted by the ABA would move forward by officially nominating "not qualified" minority or female candidates to the full Senate, while declining to move forward the nominations of "not qualified" white or male candidates. The observable implication of this selection problem is that the publicly announced ratings would appear skewed against women or minority candidates, even though there would be no bias associated with the ratings process itself.

B. George W. Bush Nominees

As noted, George W. Bush declined to allow the ABA to evaluate presumptive nominees in advance of their nominations (Gonzales 2001). Thus, during 2001–8, we have all of the ABA scores awarded, which avoids the selection problem present for other administrations. However, only 18 African American, 28 Hispanics, and 56 women were nominated to district courts during the Bush II years (table 7), which is comparatively less than other administrations. I therefore use parametric methods instead of matching. Table 8 shows results from a logit regression including race, gender, and the same battery of professional and educational characteristics. The outcome variable is whether the nominee (here, the actual nominee) received either (1) a high rating ("well qualified") or (2) a low rating ("qualified" or "not qualified"). Given the much smaller number of nominees,

11. According to Lott (2001, 46), "three potential nominees were said to have been advised that they would get a 'not qualified' rating during Bush I and nine potential nominees fell into this category for Reagan"; Bush II did not submit names for ABA "preclearance," while Obama, an exception, has had about 14 nominees whose names have not moved forward due to receiving a poor ABA mark—by far the highest share (Savage 2011). The identities of these failed presumptive nominees are strictly confidential.

Table 7. Racial/Ethnic and Gender Distribution of Judicial Nominees by President (Johnson through Obama Administrations)

President	Whites	Blacks	Hispanics	Women	*N*
Barack Obama	.70	.2	.11	.49	122
George W. Bush	.84	.06	.10	.20	283
William J. Clinton	.76	.18	.06	.28	343
George H. W. Bush	.89	.06	.05	.18	177
Ronald Reagan	.94	.02	.04	.08	302
Jimmy Carter	.78	.15	.07	.14	207
Gerald Ford	.92	.06	.02	.02	55
Richard M. Nixon	.96	.03	.01	.01	178
Lyndon B. Johnson	.92	.05	.03	.02	118

Source.—Federal Judicial Center.
Note.—Data are as of April 3, 2012.

the results are only fleetingly significant, and, in the interest of full disclosure, I also present results showing no significance. The results from the Bush II years are largely consistent with the results seen before for blacks and are suggestive of African Americans being less likely than non-Hispanic whites to receive the higher two ABA ratings. For women and for Hispanics, the effects are no longer significant, however.

C. Synthetic Candidates

The fact that George W. Bush nominated so few diverse candidates hampers the ability to extract meaningful estimates about his term. To provide additional context, I therefore artificially replicate the possible pool of confidential short-listed candidates who were dropped from consideration. Using the fact that we know the approximate number (if not the identities) of the presumptive candidates rejected by the ABA, I include in the data a set of distinct, artificially created observations designed to be the worst possible scenario for the results presented.

To create the artificial set of observations, I generated several "presumptive nominees" per president. I did so by assuming that 4% of each president's nominees were dropped at the preclearance stage. (The exception here is George W. Bush.) This is at the upper end of the actual range, which appears to be around 2%–4% (Lott 2001). The most bias would be introduced when presidents fail to move forward poorly rated whites; not moving these individuals forward while moving forward poorly rated minorities and women would result in a skewed postselection sample. Thus, I initially create an artificial sample of 62 "failed nominees" who are white, young, and poorly rated by the ABA, and I assign them those covariates least linked with higher ABA ratings (including no prior judgeships, law clerkships, or private practice experience and limited experience). Taken together, these synthetic nominees truly present the "worst case" bias for the results already presented.

Table 8. Logit Regression for US District Court Nominations Made by George W. Bush

	Model 1	Model 2	Model 3[a]
African American	−.31	−2.25**	−2.11**
	(.63)	(1.07)	(1.04)
Hispanic	−.05	−.87	−.77
	(.51)	(.80)	(.83)
Female	−.16	.04	.11
	(.38)	(.64)	(.68)
Age	.10***	.09*	.07
	(.03)	(.05)	(.05)
Top 14 law school	−.48	−.94	−.27
	(.37)	(.66)	(.78)
Private law school	.76**	.34	.37
	(.32)	(.54)	(.61)
Law clerk	−.18	−.78	−.81
	(.33)	(.63)	(.68)
Law professor	.53	.34	−.10
	(.87)	(1.59)	(1.62)
Private practice	.43	−.13	−.52
	(.47)	(.80)	(.97)
US attorney	−.001	.72	.18
	(.55)	(1.11)	(1.22)
Assistant US attorney	1.37***	1.89**	1.81**
	(.41)	(.76)	(.78)
Justice Department	−.57	−.44	−1.04
	(.65)	(1.26)	(1.52)
Public defender	.72	1.70	1.68
	(.70)	(1.55)	(1.89)
Federal magistrate	.86*	.80	.77
	(.48)	(.84)	(.90)
State judge	−.07	−.28	.13
	(.32)	(.56)	(.67)
Years of practice			.004
			(.04)
District court dummies		✓	✓
Nomination year dummies		✓	✓
N	277	277	257

Source.—Federal Judicial Center.

Note.—Outcome variable is whether the nominee received a (1) well qualified or exceptionally well qualified ABA rating vs. (2) a not qualified or qualified rating.

[a] Includes years of prenomination lawyering experience for confirmed candidates.

* $p < .1$

** $p < .05$.

*** $p < .01$.

After including them in with the original data, I reran the key analyses, which are presented in table 9. The original results are insensitive to their inclusion, particularly for African Americans and for Hispanics, for whom the relationship to high ABA scores is still negative and significant under any model specification. For women, the results are still significant once we allow the effect to vary across presidential administration

Table 9. Logit Coefficients Generated When Including 62 (4%) Generated Observations to Represent Unknown, Confidentially Dropped Presumptive Nominees

	Model 1	Model 2	Model 3
African American	−.67***	−.74***	−.76***
	(.19)	(.19)	(.23)
Hispanic	−.61***	−.75***	−.73***
	(.23)	(.23)	(.27)
Female	−.17	−.31**	−.33*
	(.15)	(.15)	(.18)
African American × female			.091
			(.44)
Hispanic × female			.16
			(.52)
Constant	−4.77***	−4.07***	−3.45***
	(.44)	(.53)	(.57)
Other controls	✓	✓	✓
President dummies		✓	✓
N	1,777	1,777	1,691

Source.—Federal Judicial Center.

Note.—Outcome variable is receiving a high ABA rating. Controls for professional experience, age, and education not shown.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

(model 2). The results are therefore not broken, even under the extreme assumptions that the only dropped presumptive nominees are poorly rated white men. In addition, incrementally increasing the number of "presumptive judges" in the artificially created set shows that the fraction of confidentially dropped nominees would have to be 7% to break the results for women, 16% to break the results for Hispanics, and 17% to break the results for African Americans—nearly twice the 7% rate recently reported for the Obama administration (Savage 2011), which is in turn about twice as high as previous administrations (Lott 2001). Moreover, given that the actual presumptive nominees are widely rumored to be minorities or women, very much unlike the underqualified white male "nominees" used here, we are left with little evidence that the results seen here are driven or even undermined by the practice of presidents secretly dropping presumptive nominees.

VI. WHY DIFFERENCES IN ABA RATINGS COULD MATTER

The above discussion highlights the fact that racial minorities and women are more likely to receive lower ratings than their white, male counterparts. I now turn to exploring the ramifications of these discrepancies. After all, if the ratings are not meaningful for policy or for political decisions, then perhaps we should be less worried that racial minorities and women appear to be less successful in achieving high ratings. Phrased differently, why should any discrepancies matter?

To examine the broader implications, I explore two issues. The first is confirmation success. Are ABA ratings somehow related to whether a candidate is more or less successful? If so, this raises questions about whether racial minorities and women could be systematically disadvantaged as they move through the confirmation process. The second is quality. Perhaps ABA ratings are important because they provide a useful signal about the potential quality of a district court candidate or his or her likely performance. To untangle this further, I examine one possible measure of judicial "quality": reversal rates (Choi et al. 2012; Epstein et al. 2013).

A. ABA Ratings and Confirmation Success

The first policy ramification I examine is the relationship between ABA ratings and confirmation success. The outcome variable here is dichotomous: whether the nominee was eventually confirmed ("1") or had his or her nomination withdrawn, killed by inaction, or rejected ("0"), as happened to 7.5% of these nominees. Table 10 shows the results of several rare-event logit regressions employing different model specifications. (I use a rare-event logit specification because of the relatively small probability of confirmation failure; however, replicating these analyses using a standard logit specification results in identical substantive inferences.) All the model specifications consistently show that high ABA ratings are the most predictive factor other than party or public defender status in determining confirmation success. Even the inclusion of controls for race (with non-Hispanic whites comprising the baseline group), gender, rank of law school attended, legal clerkships, prior judicial experience, and previous professional experience does not change the key findings, which are significant at the 1% level.[12] Neither does the inclusion of year fixed effects in columns 3 and 6—which effectively control for political factors such as Senate composition (Binder and Maltzman 2002), divided government (Binder and Maltzman 2002), increased political emphasis on lower court appointments (Hartley and Holmes 1996), or other exogenous political or legal shocks (e.g., gas prices, controversial Supreme Court decisions)—affect the main results.[13] Indeed, under all specifications, receiving a "not qualified" rating dramatically

12. Note that some of the factors might be collinear with the ratings, which would have the effect of increasing standard errors substantially. However, this does not appear to be the case for three reasons. First, although the standard errors increase on the ratings variables for table 10, cols. 1–2 and cols. 4–5 (as is expected when additional covariates are introduced), the increase in the standard errors is not so large as to suggest multicollinearity problems. Second, dropping and adding additional variables incrementally (in analyses not shown) does not substantially disrupt the coefficient estimate or standard errors of the ratings variables. Finally, having larger standard errors would primarily be a problem insofar as we would rule out effects when in fact effects exist. The two models with additional controls yield coefficient estimates on the ratings variables that are still significant. Multicollinearity would make these estimates, again which are still significant, conservative.

13. For these analyses, years in which there is perfect separation (i.e., no variance in the outcome variable) are dropped, although including them anyway results in substantively identical inferences. I also obtain identical inferences when including dummy variables for the identity of the appointing president in tandem with flexible (polynomial) controls for year (not shown).

Table 10. Rare-Events Logit Regression for US District Court Nominations between 1960 and 2002

	(1)	(2)	(3)	(4)	(5)	(6)
Qualified rating	2.38***	2.28***	3.04***			
	(.53)	(.61)	(.79)			
Well qualified rating	3.11***	2.89***	3.67***			
	(.53)	(.63)	(.80)			
High ABA rating				.87***	.73***	.80***
				(.20)	(.24)	(.28)
Age at nomination		.04*	.05**		.04**	.05**
		(.02)	(.02)		(.02)	(.02)
Female		.54	.85**		.61*	.89**
		(.32)	(.37)		(.32)	(.37)
African American		.24	.25		.24	.26
		(.39)	(.45)		(.39)	(.43)
Hispanic		.46	.29		.49	.43
		(.51)	(.57)		(.51)	(.57)
Top 14 law school		−.39	−.21		−.44*	−.31
		(.26)	(.29)		(.25)	(.28)
Law clerk		.38	.73**		.37	.70**
		(.29)	(.34)		(.29)	(.33)
Law professor		.16	.01		.11	−.01
		(.57)	(.62)		(.55)	(.60)
Private practice		.30	.45		.36	.49
		(.38)	(.41)		(.37)	(.41)
US attorney		.12	.09		.10	.10
		(.47)	(.50)		(.47)	(.49)
Assistant US attorney		.16	.19		.22	.34
		(.34)	(.37)		(.33)	(.36)
Justice Department		.33	.77		.27	.58
		(.57)	(.65)		(.56)	(.63)
Public defender		−.75*	−.42		−.77*	−.50
		(.41)	(.45)		(.40)	(.45)
Federal magistrate		.53	.56		.55	.59
		(.45)	(.50)		(.45)	(.50)
Federal bankruptcy		−1.18*	−1.10		−1.13*	−.95
		(.66)	(.83)		(.66)	(.81)
State judge		.51*	.67**		.55**	.68**
		(.27)	(.31)		(.26)	(.30)
Constant	−.59	−3.00**	−4.53***	1.73***	−1.15	−1.92
	(.60)	(1.33)	(1.60)	(.32)	(1.20)	(1.38)
President dummies	✓	✓	✓	✓	✓	✓
Year dummies			✓			✓
N	1,776	1,716	1,113	1,776	1,716	1,113

Source.—Federal Judicial Center.

Note.—Outcome variable is whether nomination succeeded (1) or failed (0). "Not qualified" is the excluded category for ABA ratings in cols. 1–3. Columns 4–6 compare high ratings vs. low ratings. In addition, cols. 3 and 6 include year fixed effects to control for other political factors, such as Senate composition or polarization.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

lowers the probability that a district court nominee will succeed. The results are also statistically significant when we dichotomize the ABA ratings into high ("exceptionally well qualified" and "well qualified") and low ("not qualified" and "qualified") ratings, as seen in table 10, columns 4–6.

Substantively, this means that poorly rated candidates are worse off in terms of confirmation success. Indeed, under the full model (table 10, col. 3) the predicted probability of confirmation success for individuals who receive "well qualified" or "qualified" ratings is high, around 95% and 96% respectively.[14] However, receiving a poor rating lowers these probabilities significantly. Those who receive only a "qualified" rating have a predictive probability of confirmation success that is around 90%, a statistically significant 5.5 percentage-point drop. Those who receive a "not qualified" rating are even worse off. For these candidates, the predicted probability of confirmation success is around 50%, a statistically significant 35 percentage-point drop from the "well qualified" candidates. Thus, avoiding a low rating from the ABA is fairly important for judicial nominees, with avoiding a "not qualified" rating, which can reduce the likelihood of confirmation success by over a third, being of particular importance.

B. Whether ABA Ratings Predict Performance

The results so far demonstrate that receiving a low ABA rating has the potential to derail a judicial candidacy. I now turn to a separate question, which concerns the utility of ABA scores. Given the effort that goes into calculating these scores and the importance assigned to them by political actors, we would expect that ABA ratings serve some useful function or signal. In the nominations context, the greatest utility would be if ABA scores somehow predict how judges will fare once on the bench. I examine one such measure of this, reversal rates. I note that reversal is not a universally agreed-on measure of judicial "quality" or "performance," which are inherently slippery concepts, and that a lively normative debate is ongoing about whether, and to what extent, judges should be held to performance standards. I also note that reversals are only one possible measure of "quality" and that others exist; Lott (2001), for example, examined lawyers' written opinions about sitting appeals court judges. When it comes to the district courts, however, there is some agreement that reversal is costly and something to be avoided (Choi et al. 2012; Epstein et al. 2013). Thus, if ABA scores are useful predictors of reversal rates, perhaps they do capture some component of "quality" or proficiency at judicial tasks.

To address these issues, I generated a second data set, this one on judges' reversal rates. Here, I relied on the Judicial Reversal Reports compiled by Westlaw, a commercial legal database. These Judicial Motion Reports display for each district judge the number of

14. All predicted probabilities are calculated via simulations that hold all other variables at their mean or mode.

cases that were appealed and how many were reversed or affirmed.[15] I was thus able to calculate for each judge his or her overall reversal rate. One important caveat is that Westlaw provides Judicial Motions Reports only from 2000 moving forward; thus, I have these data for a subset of approximately 1,100 judges, as opposed to the full 1,600, and only across cases appealed between January 2000 and July 2012.

I analyze the Westlaw data a number of different ways. For the judge reversal data, an important consideration is that the number of cases heard and appealed varies from judge to judge. This variation is to be expected and is due to different lengths of service, different jurisdictions hearing different numbers of cases, and the random fluctuations associated with case assignment (which is also usually random, conditional on district). A judge who retires in 2001 will therefore have fewer cases included in his or her judicial reversal report versus a judge who served the entirety of 2000–2012. Accordingly, an ordinary least squares specification with the simple reversal rate as the outcome would violate the basic ordinary least squares assumptions because the variance of the outcome clearly varies according to whether the judge had one case appealed or 46. I therefore take a weighted least squares approach by weighting each judge by the number of cases he or she had appealed to the designated appeals court.

Results from this model are presented in table 11. To test the specific effects of receiving the lowest rating possible, I take "not qualified" as the baseline category in columns 1–3. Even when including dummies for district court (to capture fluctuations in reversal rates across districts and circuits, cols. 2, 3, 5, and 6) and controls for professional experience (cols. 3 and 6), the results are not significant; there is no difference in reversal rates between those "not qualified" and others. Similar results are obtained when we examine judges who were rated either poorly or highly (cols. 4–6). Indeed, not only are the coefficients close to zero and statistically insignificant, but the R^2 of the simplest model (cols. 1 and 4) is close to zero as well; what predictive power we have comes not from the ABA ratings but from the addition of district dummies and additional controls (cols. 2, 3, 5, and 6).

A second methodological challenge is, however, that we only have case-reversal data on judges who were confirmed by the Senate; thus, there are no data on those whose nominations were withdrawn, rejected, or killed due to Senate inaction—a selection process affected by the ABA ratings themselves (see above). To address this nonrandom selection issue, I take a missing-data-style approach. First, I fit a regression that predicts the probability of selection into the sample (i.e., confirmation). Second, I then focus on the confirmed judges and fit a weighted linear model that takes as the key outcome the judges' reversal rates and includes along with the weights the predicted probability of inclusion into the sample. This is effectively making an missing-at-random assumption,

15. This is reported by Westlaw as the total number of cases "affirmed" and does not include cases that were affirmed in part and reversed in part.

Table 11. Weighted Ordinary Least Squares Regression

	(1)	(2)	(3)	(4)	(5)	(6)
Qualified	−.01	.004	.02			
	(.04)	(.03)	(.03)			
Well qualified	−.01	.002	.01			
	(.04)	(.03)	(.03)			
Exceptionally well qualified	−.02	−.001	.01			
	(.05)	(.03)	(.03)			
High ABA rating (yes or no)				−.003	−.002	−.01
				(.01)	(.004)	(.004)
Constant	.33***	.85**	.79**	.32***	.85**	.80**
	(.04)	(.33)	(.33)	(.004)	(.33)	(.33)
Professional controls			✓			✓
Nomination year dummies		✓	✓		✓	✓
District court dummies		✓	✓		✓	✓
N	1,131	1,131	1,121	1,131	1,131	1,121
R^2	.0004	.66	.68	.0003	.66	.68

Sources.—Westlaw Judicial Motion Reports, Federal Judicial Center.

Note.—Outcome is a judge's reversal rate for cases decided on appeal between 2000 and 2012. "Not qualified" is the excluded ABA ratings category in cols. 1–3. Professional controls include all educational, professional, and practical experience variables used in other analyses. All observations are weighted by the number of cases appealed per judge.

** $p < .05$.

*** $p < .01$.

which is that conditional on the observables (e.g., ABA rating), the reversal rate of the nominee should not depend on being confirmed. That is, conditional on all observables, the distribution of reversal rates among judges confirmed should be the same among those not confirmed.

These results are reported in table 12 and confirm the results presented above. Under no model specification are ABA ratings predictive of reversal rates. This is the case when we compare the "not qualified" rating to the three others (cols. 1 and 2) and also when we compare poorly rated judges to highly rated ones (cols. 3 and 4). The addition or omission of other controls (for circuit or for professional experience) does not change the inferences. Taken together, the analyses point to two conclusions: first, ABA ratings do little to predict a judge's ultimate reversal rate, and, second, "not qualified" judges are no more likely to be overturned than are their higher-rated peers.

C. Case-Level Data

The Westlaw data are aggregated at the judge level; however, reversal is probably likely to vary according to case-specific characteristics (Epstein et al. 2013). Thus, to guard against these findings being data or model dependent, I also present results from case-level data collated by Songer (2007) and Kuersten and Haire (2011). These data represent approximately 9,000 randomly selected published cases from the US courts of appeals up to 2002, which allows me to investigate the relative importance of ABA ratings

Table 12. Weighted Ordinary Least Squares Regression with Selection Correction

	(1)	(2)	(3)	(4)
Qualified	.01	.02		
	(.02)	(.02)		
Well qualified	.004	.01		
	(.02)	(.02)		
Exceptionally well qualified	−.0000	.01		
	(.03)	(.03)		
High ABA rating (yes or no)			−.003	−.01
			(.004)	(.004)
Constant	.85**	.79**	.85**	.79**
	(.34)	(.33)	(.34)	(.33)
Professional controls		✓		✓
Nomination year dummies	✓	✓	✓	✓
District court dummies	✓	✓	✓	✓
N	1,114	1,121	1,114	1,121
R^2	.66	.68	.66	.68

Sources.—Westlaw Judicial Motion Reports, Federal Judicial Center.

Note.—Outcome is a judge's reversal rate for cases decided on appeal between 2000 and 2012. "Not qualified" is the excluded ABA ratings category in cols. 1 and 2. Professional controls include all educational, professional, and practical experience variables used in other analyses. All observations are weighted by the number of cases appealed per judge.

** $p < .05$.

when examining individual cases.[16] Along with the appeals court votes, this combined data set includes coding on case issue area, parties, amici, and other case-level factors. I combine these data with two data sets containing judicial common space scores, one at the appeals level (Epstein et al. 2007) and the other at the district court level (Boyd 2011). This allows me to control for the ideological distance between the district and the appeals courts, which could potentially affect reversal rates (Choi et al. 2012).

I analyzed these case-level data by looking at whether any one case was reversed (coded as "1") or affirmed ("0").[17] I also include (1) dummies for the district court, (2) dummies for district judge race and gender, (3) district judge professional and educational characteristics, (4) dummy variables for the year the appeals court decided the case, (5) an indicator for whether the district judge had taken senior status at the time of the appeals court decision, (6) controls for the number of parties listed as appellants or respondents, (7) dummy variables for the appointing president of the lower-court judge, and

16. The analyses using the Westlaw data presented in Sec. VI.B include both published and unpublished cases. Because the perceived "quality" of a district judge might affect a panel's decision whether to publish a case, it is important to include both published and unpublished cases so as to avoid any potential biases. This analysis therefore complements, rather than replaces, the reversal rate analysis.

17. Specifically, I use Kuersten and Haire's (2011) coding and denote a case as affirmed if it was affirmed or affirmed in part. I denote that a case is reversed if it was reversed or vacated. I err on the side of being conservative; if any case was affirmed or upheld in part, it is considered "affirmed."

(8) controls for case characteristics, which include the number of amici involved,[18] the case's procedural posture,[19] the case issue area,[20] and the identity of the parties.[21] The last model also includes (9) the absolute value of the difference between the judicial common space scores for the district judge and for the median member of the appeals panel so as to capture whether the lack of results is driven simply by ideological distance, as suggested by Choi et al. (2012). In addition, because each judge appears multiple times in the data, I nclude judge-specific random effects; not including these would treat cases decided by the same judge as independent, which would have the effect of underestimating the standard errors.

Results from this analysis are presented in tables 13 and 14. Because of the large number of potentially relevant control variables, I add them incrementally. Across all model specifications, and across all possible ways of stylizing the key explanatory variable, the results are the same. There appears to be no relationship between a district judge's ABA rating and the probability that a case written by that judge will be reversed on appeal. Judges rated "not qualified" are no more likely to be reversed on appeal than judges who are rated "well qualified" or "exceptionally well qualified" (table 13). Neither are judges rated "not qualified" or "qualified" more likely to be overturned than those rated "well qualified" or "exceptionally well qualified" (table 13). These analyses provide reassurance that other factors—such as parties involved, case characteristics, procedural posture, or ideological differences—are not driving the results.

VII. WHY ARE JUDICIAL NOMINEES RATED?

The conclusions of this article are threefold. First, my findings suggest that minority and female judicial candidates systematically receive lower qualification ratings from the ABA. This is the case both a priori and also when using matching or other controls to compare candidates who are similar across key professional, educational, and political characteristics. The results also appear robust to potential omitted variables and to possible selection bias that occurs when presidents privately decline to pursue certain

18. Per Kuersten and Haire's (2011) coding, this measure varies between zero briefs filed and seven briefs filed. Separate dummy variables represent cases with eight or more briefs or cases for which the number of briefs is impossible to ascertain. This variable is included in the analysis as a categorical variable.

19. Per Kuersten and Haire's (2011) coding, this is a categorical variable that takes on discrete values for whether the case arose on appeal from (1) a jury or bench trial, (2) an injunction, (3) a motion for summary judgment, (4) a guilty plea, (5) a dismissal, (6) postjudgment orders, (7) postsettlement orders, (8) an interlocutory appeal, (9) a writ of mandamus, (10) other, (11) miscellaneous, or (12) not applicable.

20. This includes dummy variables for whether the case involved (1) criminal law, (2) civil rights, (3) First Amendment, (4) due process, (5) privacy, (6) labor relations, (7) economic activity and regulation or was miscellaneous or not ascertained (Kuersten and Haire 2011).

21. This was operationalized as indicator variables for whether the appellant or the respondent was a natural person, a business, the federal government, or a state, as coded by Kuersten and Haire (2011).

Table 13. Logit Regression Results from Randomly Selected Court of Appeals Cases, Coded by Songer (2007) and Kuersten and Haire (2011)

	(1)	(2)	(3)	(4)
Qualified	.26	.08	−.05	−.39
	(.34)	(.36)	(.38)	(.42)
Well qualified	.16	−.02	−.15	−.49
	(.34)	(.36)	(.38)	(.42)
Exceptionally well qualified	.26	.09	−.15	−.51
	(.36)	(.38)	(.41)	(.45)
Age		.003	.002	.004
		(.01)	(.01)	(.01)
African American		.16	.24*	.27**
		(.11)	(.13)	(.13)
Hispanic		−.17	−.14	−.11
		(.20)	(.22)	(.22)
Female		.16	.12	.17
		(.11)	(.12)	(.12)
Top 14 law school		−.04	−.09	−.09
		(.07)	(.07)	(.08)
Private law school		.01	−.005	.01
		(.07)	(.08)	(.08)
Law clerk		−.02	.01	−.04
		(.09)	(.10)	(.10)
Law professor		−.02	.01	.02
		(.12)	(.13)	(.13)
Private practice		−.01	.03	.05
		(.14)	(.16)	(.16)
US attorney		.06	.09	.08
		(.10)	(.11)	(.12)
Assistant US attorney		−.003	.08	.10
		(.08)	(.09)	(.09)
Justice Department		.03	−.09	−.10
		(.14)	(.16)	(.16)
Public defender		.36	.19	.12
		(.22)	(.24)	(.25)
Federal magistrate		.31**	.21	.26
		(.15)	(.17)	(.17)
Federal bankruptcy		.64***	.70***	.64**
		(.23)	(.25)	(.26)
State judge		.17**	.26***	.26***
		(.07)	(.08)	(.08)
Senior status			−.07	−.08
			(.12)	(.10)
Total parties			−.01	−.01
			(.01)	(.01)
Ideology divergence				.16
				(.12)
Constant	−.73**	−14.22	−29.03	1.83
	(.35)	(1,459.97)	(2,931.19)	(1,783.71)
Judge random effects	✓	✓	✓	✓
District court dummies	✓	✓	✓	✓

Table 13. (*Continued*)

	(1)	(2)	(3)	(4)
Year dummies		✓	✓	✓
Appointing president dummies		✓	✓	
Case complexity controls			✓	✓
N	7,698	7,654	6,720	6,508

Sources.—Federal Judicial Center, Epstein et al. (2007), Songer (2007), Boyd (2011), and Kuersten and Haire (2011).

Note.—Outcome variable is whether case was reversed (1) or affirmed (0). "Not qualified" is the excluded category for ABA ratings. Model 4 includes absolute difference between appeals panel and district court ideology.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

nominees. Indeed, the results suggest that this difference persists even when we take into account the fact that African Americans, Hispanics, and women might be drawn to different career trajectories (e.g., public defender or government work) or have different years of trial or practice experience.

Second, turning to the question of why these discrepancies could be important, I show that ABA ratings are a significant predictor of whether a judicial nominee will be confirmed, with "not qualified" rated nominees being much more likely to fail—even when controlling for other personal and professional characteristics. Thus, ABA ratings matter, and avoiding very low ratings is paramount for judicial nominees. Finally, I find that ABA ratings are not predictive of judges' ultimate performance once they are confirmed. Indeed, nominees designated as "not qualified" or "qualified" have reversal rates that differ little from those awarded the stellar "exceptionally well qualified" and "well qualified" ratings. This fact is surprising given that the ABA ostensibly takes into account those aspects that would make for a strong judicial career—both objective criteria like law school attended and also more qualitative criteria such as "temperament," "competence," and "integrity."

Why should minorities and women receive lower ratings? One way to try to understand these puzzling results is that the law is a prestige-oriented profession—one driven by high-status accomplishments and the general appearance of success. To this extent, it is not surprising that rank of law school, assistant US attorney experience, previous legal clerkships, and success in private practice are predictive of the kind of ABA rating a nominee will receive. However, in instances where prestige, power, and appearances matter, we might also not be surprised that women, minorities, and other individuals who have traditionally held less prestigious positions might be systematically disadvantaged. This is particularly the case once we consider the fact that the ABA itself uses criteria through which social biases themselves may be perpetrated. For example, "integrity" and "judicial temperament," two of the ABA's criteria, are highly subjective standards, which, considered separately, could easily incorporate certain biases in favor of

Table 14. Logit Regression Results from Randomly Selected Court of Appeals Cases, Coded by Songer (2007) and Kuersten and Haire (2011)

	(1)	(2)	(3)	(4)
High ABA rating (yes or no)	−.09	−.09	−.11	−.10
	(.05)	(.06)	(.07)	(.07)
Age		.004	.002	.004
		(.01)	(.01)	(.01)
African American		.16	.24*	.27**
		(.11)	(.13)	(.13)
Hispanic		−.16	−.14	−.12
		(.20)	(.22)	(.22)
Female		.16	.12	.17
		(.11)	(.12)	(.12)
Top 14 law school		−.03	−.09	−.09
		(.07)	(.07)	(.08)
Private law school		.01	−.004	.02
		(.07)	(.08)	(.08)
Law clerk		−.02	.01	−.04
		(.09)	(.10)	(.10)
Law professor		−.02	.01	.03
		(.12)	(.13)	(.13)
Private practice		−.01	.03	.05
		(.14)	(.16)	(.16)
US attorney		.06	.09	.08
		(.10)	(.11)	(.11)
Assistant US attorney		−.01	.08	.10
		(.08)	(.09)	(.09)
Justice Department		.04	−.09	−.10
		(.14)	(.16)	(.16)
Public defender		.36	.19	.12
		(.22)	(.24)	(.25)
Federal magistrate		.31**	.21	.26
		(.15)	(.17)	(.17)
Federal bankruptcy		.64***	.70***	.64**
		(.23)	(.25)	(.26)
State judge		.17**	.26***	.25***
		(.07)	(.08)	(.08)
Senior status			−.07	−.08
			(.12)	(.10)
Total parties			−.01	−.005
			(.01)	(.01)
Ideology divergence				.16
				(.12)
Constant	−.46***	−14.16	−29.08	1.45
Judge random effects	✓	✓	✓	✓
District court dummies	✓	✓	✓	✓
Year dummies		✓	✓	✓
Appointing president dummies		✓	✓	
Case complexity controls				✓
N	7,698	7,654	6,720	6,508

Sources.—Federal Judicial Center, Epstein et al. (2007), Songer (2007), Boyd (2011), and Kuersten and Haire (2011).

Note.—Outcome variable is whether case was reversed (1) or affirmed (0). Key explanatory variable is whether judge was rated highly (exceptionally well qualified or well qualified) by the ABA. Model 4 includes absolute difference between appeals panel and district court ideology.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

whites and men (e.g., the group that society has historically decided possess judicial "integrity" or "temperament"). This is not to say that the ABA is engaging in discriminatory practices, but it is to say that we cannot rule out the possibility of implicit bias against these sorts of nominees, which would perhaps be unsurprising given the wealth of other studies finding implicit biases at high-level organizations (Bielby and Baron 1986; Fernandez et al. 2000; Castilla 2008). Having a ratings process that is more transparent and more candid about the exact criteria used might help shed light on the roots of these stubborn discrepancies.

Although this analysis resolves some questions, it leaves others open. Indeed, this analysis has shown that an increasingly large segment of nominees appears to systematically receive lower ratings; at the same time, the ratings themselves do little to predict whether these judges will be better or worse in terms of reversal rates. Why, if ABA ratings do so little to predict judicial performance, do senators and presidents continue to rely on them? Once possible explanation is that political actors simply do not know that ABA ratings do little to predict judicial performance. But perhaps another answer is that political actors rely on these ratings for reasons unrelated to the appointments process. For example, although the ABA is avowedly nonpartisan and makes no campaign contributions itself, the legal industry is the third-largest source of campaign and political contributions. By some estimates law firms and lawyers made some $390 million in campaign donations in 2008, with approximately 70% of these donations going to Democratic candidates. Barack Obama himself has been the recipient of $57 million from the legal industry, compared with $11 million to George W. Bush and $8 million to Mitt Romney. The legal industry spends an additional $50 million per year on lobbying efforts. The scale and largesse of these gifts is hard to ignore; given the low cost (to political actors) associated with agreeing to vet candidates, and with the huge risk of alienating or antagonizing a wealthy and giving constituency, it makes sense that political actors continue to seek out ABA ratings. Admittedly, for Republicans, the risks are comparatively less—which might explain the willingness of some Republicans to speak out against the use of ABA ratings and the choice of one administration to eschew them altogether. However, for Democrats, they must proceed with caution; continuing to incorporate ABA ratings into their calculations risks alienating another constituency, minorities and women. Whether the benefits associated with using these kinds of ratings will continue to outweigh their potential political risks remains to be seen.

REFERENCES

ABA (American Bar Association). 2009. "Standing Committee on the Federal Judiciary: What It Is and How It Works." ABA, Chicago.

Allison, Garland W. 1996. "Delay in Senate Confirmation of Federal Judicial Nominees." *Judicature* 80 (1): 8–15.

Banks, R. Richard, Jennifer L. Eberhardt, and Lee Ross. 2006. "Discrimination and Implicit Bias in a Racially Unequal Society." *California Law Review* 94 (4): 1169–90.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94 (4): 991–1013.

Bielby, William T., and James N. Baron. 1986. "Men and Women at Work: Sex Segregation and Statistical Discrimination." *American Journal of Sociology* 91 (4): 759–99.

Binder, Sarah A., and Forrest Maltzman. 2002. "Senatorial Delay in Confirming Federal Judges." *American Journal of Political Science* 46 (1): 190–99.

Boyd, Christina L. 2011. "Federal District Court Judge Ideology Data." Author's web page, University of Georgia. http://cLboyd.net/ideology.html.

Butler, Daniel M., and David E. Broockman. 2011. "Do Politicians Racially Discriminate against Constituents? A Field Experiment on State Legislators." *American Journal of Political Science* 55 (3): 463–77.

Castilla, Emilio J. 2008. "Gender, Race, and Meritocracy in Organizational Careers." *American Journal of Sociology* 113 (6): 1479–1526.

Choi, Stephen J., Mitu Gulati, and Eric A. Posner. 2012. "What Do Federal District Judges Want? An Analysis of Publications, Citations, and Reversals." *Journal of Law, Economics, and Organization* 28 (3): 518–49.

De Rohan Barondes, Royce. 2010. "ABA Judicial Ratings as the 'Gold Standard' or Fool's Gold: Federal District Judge Ratings and Reversals." Working Paper, School of Law, University of Missouri–Columbia.

Epstein, Lee, William M. Landes, and Richard A. Posner. 2013. *The Behavior of Federal Judges: A Theoretical and Empirical Study of Rational Choice*. Cambridge, MA: Harvard University Press.

Epstein, Lee, Andrew D. Martin, Jeffrey A. Segal, and Chad Westerland. 2007. "The Judicial Common Space." *Journal of Law, Economics, and Organization* 23 (2): 303–25.

Fernandez, Roberto M., Emilio J. Castilla, and Paul Moore. 2000. "Social Capital at Work: Networks and Employment at a Phone Center." *American Journal of Sociology* 105 (5): 1288–1356.

Giles, M. W., V. A. Hettinger, and T. Peppers. 2001. "Picking Federal Judges: A Note on Policy and Partisan Selection Agendas." *Political Research Quarterly* 54 (3): 623–41.

Goldman, Sheldon. 1997. *Picking Federal Judges: Lower Court Selection from Roosevelt through Reagan*. New Haven, CT: Yale University Press.

Gonzales, Alberto. 2001. "Letter to the ABA from Al Gonzales." White House website, March 22. http://georgewbush-whitehouse.archives.gov/news/releases/2001/03/20010322-5.html.

Greenwald, Anthony G., and Linda Hamilton Krieger. 2006. "Implicit Bias: Scientific Foundations." *California Law Review* 94 (4): 945–67.

Haire, Susan. 2001. "Rating the Ratings of the American Bar Association Standing Committee on Federal Judiciary." *Justice System Journal* 22 (1): 1–17.

Hartley, Roger E., and Lisa M. Holmes. 1996. "Increasing Senate Scrutiny of Lower Federal Court Nominees." *Judicature* 80:274–78.

Ho, Daniel E. 2005. "Why Affirmative Action Does Not Cause Black Students to Fail the Bar." *Yale Law Journal* 114 (8): 1997–2005.

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15 (3): 199–236.

Iacus, Stefano M., Gary King, and Giuseppe Porro. 2009. "CEM: Software for Coarsened Exact Matching." Software documentation, Harvard University. http://j.mp/oGIFsf.

———. 2011. "Multivariate Matching Methods That Are Monotonic Imbalance Bounding." *Journal of the American Statistical Association* 106 (493): 345–61.

Just the Beginning Foundation. 2012. "Integration of the Federal Judiciary." Just the Beginning Foundation, Chicago. http://www.jtbf.org.

Kang, Jerry. 2005. "Trojan Horses of Race." *Harvard Law Review* 118 (5): 1489–1593.

Keele, Luke. 2010. "An Overview of rbounds: An R Package for Rosenbaum Bounds Sensitivity Analysis with Matched Data." Software documentation, Pennsylvania State University. http://www.personal.psu.edu/ljk20/rbounds%20vignette.pdf.

Koenig, Dustin. 2012. "Bias in the Bar: ABA Ratings and Federal Judicial Nominees from 1976–2000." *Judicature* 95 (4): 188–95.

Krieger, Nancy, Dana Carney, Katie Lancaster, Pamela D. Waterman, Anna Kosheleva, and Mahzarin Banaji. 2010. "Combining Explicit and Implicit Measures of Racial Discrimination in Health Research." *American Journal of Public Health* 100 (8): 1485–92.

Kuersten, Ashlyn K., and Susan B. Haire. 2011. "Update to the Appeals Courts Database, 1997–2002." Western Michigan University. http://www.wmich.edu/nsf-coa.

Lindgren, J. T. 2001. "Examining the American Bar Association's Ratings of Nominees to the US Courts of Appeals for Political Bias, 1989–2000." Public Law and Legal Theory Paper no. 37, School of Law, Northwestern University.

Lott, John R., Jr. 2001. "The American Bar Association, Judicial Ratings, and Political Bias." *Journal of Law and Politics* 17:41.

———. 2005. "The Judicial Confirmation Process: The Difficulty with Being Smart." *Journal of Empirical Legal Studies* 2 (3): 407–47.

Nixon, David C., and David L. Goss. 2001. "Confirmation Delay for Vacancies on the Circuit Courts of Appeals." *American Politics Research* 29 (3): 246–74.

Raben, Robert. 2011. "The ABA Ratings and Minority Nominees: Shedding Light on Disparate Impact." Hispanic National Bar Association, Washington, DC. http://www.hnba.com/category/blogs/op-ed/.

Rosenbaum, Paul R. 2002. *Observational Studies*. New York: Springer.

Saks, Michael J., and Neil Vidmar. 2001. "Flawed Search for Bias in the American Bar Association's Ratings of Prospective Judicial Nominees: A Critique of the Lindgren Study." *Journal of Law and Politics* 17:219–54.

Samuel, R. B. 2006. "The Structural Turn and the Limits of Antidiscrimination Law." *California Law Review* 1:5–10.

Sander, Richard H. 2004. "A Systemic Analysis of Affirmative Action in American Law Schools." *Stanford Law Review* 57 (2): 367–483.

Savage, Charles. 2011. "Ratings Shrink President's List for Judgeships." *New York Times*, November 22.

Scherer, Nancy, Brandon L. Bartels, and Amy Steigerwalt. 2008. "Sounding the Fire Alarm: The Role of Interest Groups in the Lower Federal Court Confirmation Process." *Journal of Politics* 70 (4): 1026–39.

Smelcer, Susan Navarro, Amy Steigerwalt, and Richard L. Vining. 2012. "Bias and the Bar: Evaluating the ABA Ratings of Federal Judicial Nominees." *Political Research Quarterly* 65 (4): 827–40.

Songer, Donald B. 2007. "The United States Courts of Appeals Database, 1925–1996." Judicial Research Initiative, University of South Carolina. http://artsandsciences.sc.edu/poli/juri/appct.htm.

Stratmann, Thomas, and Jared Garner. 2004. "Judicial Selection: Politics, Biases, and Constituency Demands." *Public Choice* 118 (3–4): 251–70.

Tetreault, S. 2010. "Reid Criticizes Lawyers Group." *Law Vegas Review-Journal*, February 12.