Is Justice Really Blind?
Race and Appellate Review in U.S. Courts∗†

Maya Sen‡
msen@ur.rochester.edu

Abstract

In this paper, I use two new data sets to demonstrate that black federal judges are consistently overturned on appeal more often than white judges. The racial gap is robust and persists after taking into account previous professional and judicial experience, educational backgrounds, qualification ratings assigned by the American Bar Association, and differences in partisanship. This study is the first to explore how higher-court judges evaluate opinions written by judges of color, and it has clear implications: despite attempts to make judiciary more reflective of the general population, racial disparities within the legal system continue to persist.

Word Count: 8,496

∗I am grateful to Paul Brace, Matthew Blackwell, Adam Glynn, Anna Harvey, Jennifer Hochschild, Jonathan Kastellec, Gary King, Bethany Lacina, Clayton Nall, Richard Nielsen, Kevin Quinn, Joseph Ura, and participants at the Society for Political Methodology summer meeting, the Midwest Political Science Association Annual Meeting, the Harvard Applied Statistics Workshop, and the Conference on Empirical Legal Studies for helpful feedback. Thanks also to the Federal Judicial Center and especially to the Harvard Law Library and Westlaw staff.

†An earlier draft of this paper won the 2012 Best Graduate Student Paper Prize, American Political Science Association Law & Courts Section.

‡Assistant Professor, Department of Political Science, University of Rochester, Harkness Hall 322, Rochester NY 14627 (http://mayasen.org).
1 Introduction

In 1961, Illinois state judge James Parsons was at his summer home when he got a call that changed his life. The call was from John F. Kennedy, and over the course of the call, Kennedy asked Parsons if he would accept a federal judgeship at the U.S. District Court for the Northern District of Illinois. As Parsons later recalled, “I said, ‘As a former naval officer, aye, aye sir,’ And he said, ‘Carry on.’” The significance of this conversation – an otherwise routine exchange between a President and a potential judicial nominee – was that Parsons was black, and his investiture made him the first African American appointed to the U.S. District Courts.

Thanks to jurists like James Parsons, numerous men and women of color now occupy roles in the upper echelons of the judiciary, not just in state and federal courts, but also in other countries and at the international level. And while social scientists have an increasing understanding of how characteristics such as race influence decision making, less well understood is how the legal system has incorporated these actors – that is, how the decisions rendered by minority and women judges have been evaluated by higher courts, whether they have been treated on equal footing, and how influential they have been. On the one hand, the increased appointment of women and minorities serves to make the judiciary more reflective of the population it serves. On the other, if these judges are more likely to be overturned, then we must consider whether more needs to be done to achieve the goals of descriptive representation in the courts.

In this paper, I offer the first study detailing how higher-court judges evaluate opinions written by minority judges. I leverage two new data sets – one on the characteristics of approximately 1,100 federal district judges and the other on their corresponding appeal and reversal rates on cases decided between 2000 and 2012. By then controlling for measures of judge qualifications (including ratings awarded by the American Bar Association), profes-
sional and judicial experience, caliber of legal training, and partisanship, I find that cases decided by African-American lower court judges are up to 7% more likely to be overturned than are cases written by similar white judges, resulting in up to 2,000 additional black-authored federal cases being overturned in this time period. This racial gap is significant, robust, and appears particularly strong among judges appointed by Democratic presidents.

Although I explore several possible explanations behind this finding, the underlying mechanism is not straightforward. One possibility is that the racial gap is explained by differences in ideological views (perhaps because black judges are more liberal and are therefore overturned more). To test this, I examine the composition of reviewing appeals panels using an existing data set of randomly selected published appeals cases compiled by Songer, Kuersten and Haire (2007). I find that the difference between black and white judges in terms of reversal does not vary across more or less conservative higher courts. Neither are the results driven by distinctive voting by African Americans on civil rights or affirmative action issues, on which previous scholarship has suggested differentiated voting patterns. These results suggest that something more than simple ideological differences are at play; a more likely explanation is that the racial gap is driven by an amalgam factors, including the possibility of implicit biases by higher courts. However, the results are clear: one of the most significant factors predicting whether a judge will be reversed is, surprisingly, his or her race.

This paper proceeds as follows. I discuss theories linking race, decision making, and reversal in Parts 2 and 3. In Part 4, I discuss the data, which are two new data sets and one existing data set on U.S. district judges. Part 5 presents the core results: black judges are indeed more likely to be reversed than white judges, and the difference is statistically significant under different model specifications. In Parts 7–10, I discuss possible explanations behind this finding, which are (1) the possibility that black lower-court judges have more, or different, kinds of cases appealed to higher courts, (2) possible differences in professional experience, (3) differences in ideology, (4) differences in voting patterns in certain issue areas,
and (5) implicit racial bias by higher courts. I conclude in Part 11 with a discussion of the limitations and implications of this research.

2 Race and Judicial Decision-Making

Ever since Jimmy Carter began nominating women and minority judges in large numbers, scholars have stressed their potential importance and focused closely on their impact. On the one hand, there is a view that simply having a diverse bench might be normatively desirable (Pitkin, 1967) and that it has the potential to increase the institutional legitimacy of the courts (Scherer and Curry, 2010). Another view is that descriptive representation can also be instrumentally important by bringing important viewpoints that might otherwise be unshared. Descriptive representation can therefore often (although not always does) result in substantive representation (Krislov, 1974).

Empirical studies on this topic have mostly focused on this second question: Whether women and minority judges decide cases differently than their white male counterparts. For the most part, the answer to this question has been yes, but that they do so in the context of substantively salient issues. For example, for the minority judges who are the focus of this study,1 Kastellec (2012) finds that black judges are more likely to vote in favor of affirmative action policies and that having blacks on an appellate panel changes how that panel votes. Cox and Miles (2008) similarly find that the addition of a black judge to a panel increases the likelihood that it will find a violation of the Voting Rights Act. Other studies find difference in voting in related civil rights areas. For example, Pinello (2003) finds that black judges are more likely to side with LGBT claimants than white judges, and Martin and Pyle

---

1A similar literature addresses differences in voting by male and female judges. This scholarship suggests that there are differences in the way that male and female judges vote, but only in the context of gender-salient cases – e.g., sex discrimination (Boyd, Epstein and Martin, 2010; Baldez, Epstein and Martin, 2006; Peresie, 2005; Massie, Johnson and Gubala, 2002; Segal and Spaeth, 2002; Crowe, 1999). Others have found little or no effects associated with a judge’s gender (Manning, Carroll and Carp, 2004; Kulik, Perry and Pepper, 2003; Ashenfelter, Eisenberg and Schwab, 1995). With regard to reversal rates, I find no differences between male and female judges.
find that black judges are more likely to rule in a liberal direction in discrimination and gender-related cases. (On this last point, however, Segal (2000) finds evidence to the contrary.) Importantly, a number of these studies have found this difference between white and black judges when the race of the parties is a salient issue. In the criminal context, Scherer (2004), finds that black judges are more likely to accept black defendants’ claims of police misconduct, while Welch, Combs and Gruhl (1988) and Gottschall (1983) find that black judges are more lenient with black defendants than white judges (but see Spohn (1990), which finds no differences). I also note that a number of studies have found no differences across other legal areas (Walker and Barrow, 1985; Gottschall, 1983).

A thread running through this literature is that differences in voting are due to different personal and professional experiences, which in turn inform legal views. African Americans on the bench tend to be clustered in certain (oftentimes urban) districts, with a greater share having experience as public defenders, government lawyers, and law professors – which are thought by some to be less desirable careers (Savage, 2011). In addition, a number of black judges have historically come from the trenches of the civil rights movement (e.g., Thurgood Marshall, Constance Baker Motley, Matthew Perry), or possibly have more experience with the criminal justice system (Scherer, 2004). These different experiences could introduce or reinforce distinct attitudes about affirmative action, civil rights, and voting rights. A number of these studies examining voting differences between black and white judges do attempt to control for different political views (by way of judicial common space scores or other proxies for judicial ideology); however, that differences between black and white judges persist despite controlling for such measures suggests that blacks’ voting transcends measurements of political ideology, particularly on civil rights issues. That is, the views of black judges differ from the views of otherwise similarly liberal whites.

However, differences in voting is only half of the story, at least regarding substantive representation. Although we know that minority judges vote differently once on the bench,
we have little sense of how they are perceived or evaluated – that is, what kind of impact these judges make. The question is key for understanding descriptive representation’s impact: after all, if these judges have diminished impact due to consistent reversal or lessened influence, then their substantive impact will be lessened as well. Underlying this concern is the possibility of implicit biases against minority actors, especially in the form of more appeals and increased reversal. In this regard, a number of studies have demonstrated implicit biases against African Americans in a host of settings, including high-level business organizations (Castilla, 2008; Bielby and Baron, 1986; Fernandez, Castilla and Moore, 2000), law (Greenwald and Krieger, 2006; Banks, Eberhardt and Ross, 2006; Bagenstost, 2006; Kang, 2004), public health (Krieger et al., 2010), academia (Ginther et al., 2011), employment (Bertrand and Mullainathan, 2004; Fryer and Levitt, 2004a), housing (Yinger, 1986), and even the halls of Congress (Butler and Broockman, 2011). That the same could apply to the judiciary may be problematic, but perhaps unsurprising.

Despite substantial literature in other fields, the literature of how the legal system incorporates (non-criminal defendant) minorities is limited. Some insight comes from state-level analyses, specifically attempts to quantify judicial performance in anticipation of judicial elections. Such judicial evaluations have been implemented in 19 states and usually involve surveys of local attorneys about judicial performance (Pelander, 1998; Gill, Lazos and Waters, 2010). With regard to minority judges, Gill, Lazos and Waters (2010) have found that attorney surveys routinely award lower scores to women and minorities, even after controlling for experience and reversal rates. At the federal level, no study has looked at the comparative performance of minority or women judges, or at how often these judges are overturned by higher courts. However, perhaps the only measure of judicial “quality” comes in the form of ratings awarded by the American Bar Association (ABA); here, black judges have been shown to receive lower ratings by some studies (Lott, 2001), but not others (Smelcer, Steigerwalt and Jr, 2011). In addition, within public opinion, Scherer and Curry (2010) finds
that many perceive black judges to be more liberal.

3 Theory of Race and Judicial Reversal

In this study, I explore the judicial analogy to those outcomes explored in other implicit bias studies: reversal by higher courts. Although no study has looked at how reversal rates vary from judge to judge, the existing literature informs the hypotheses regarding race and reversal. Indeed, the fact that black judges have been shown to vote differently than white judges (Kastellec, 2012; Cox and Miles, 2008) has suggestive implications: if black judges vote differently than white judges, it would not be surprising that black lower-court judges have cases appealed at higher rates, and then are also more likely to have those cases reversed – particularly if reviewed by white appeals judges without the same political or legal inclinations. Taken in tandem with implicit biases against African Americans in other fields, and also within law and the legal system, this would suggest the first hypothesis: African-American judges will be more likely to be reversed on appeal than white judges. This forms the core inquiry of this study.

However, addressing this hypothesis without rigorously addressing possible explanations behind such an effect would be unsatisfying. I therefore consider several possible explanations. The first concerns the gatekeeping issue highlighted above – the very decision to appeal. Because litigants have discretion in choosing to appeal, not all cases are appealed, and this could vary by the race of the lower-court judge in ways that complicate any findings. Here, I consider two possibilities. First, because the existing literature suggests that black judges vote differently than white judges on certain issues, and because most appellate panels are comprised of all-white panels, losing litigants in such cases may have some incentives to appeal and secure a reversal. A second possibility is grounded in the fact that practicing attorneys are known to have lower opinions of minority judges (at least at the state level, Gill, Lazos and Waters (2010)). Thus, attorneys might view those black judges’ opinions
with more skepticism and will be more inclined to appeal them. These two mechanisms have a clear observable implication, which is my second hypothesis: opinions written by African Americans, particularly those on civil rights issues, will on average be more likely to be appealed than those written by white judges. Finding such a difference may suggest that discrepancies in reversal rates stem more from the nature of cases appealed, rather than any kind of bias by appeals panels.

Parallel narratives would also have observable implications at the reversal stage (i.e., conditional on appeal). That black judges have different, perhaps stronger or more resolute beliefs about affirmative action or voting rights (Kastellec, 2012; Cox and Miles, 2008) suggests that black judges would be reversed more in these issue areas, and in these areas only. Contrariwise, if the driving narrative was one where black judges are reversed more because they are more liberal “across the board,” then we would expect to see African American judges being reversed more frequently across a wider swath of legal topics. However, black judges would be more likely to be reversed by more conservative 3-judge appeals panels, i.e., panels with two or more Republican appointees. Thus, another explanatory hypothesis is that the “black judges” effect should increase in (1) issue areas involving civil rights and/or (2) when appeals are heard by more conservative appeals panels. Finding any of these effects would suggest that disparate reversal rates stem from differences in voting behavior rather than other causes.

There are two further explanations, both of which raise more troubling normative implications. The first is that black judges could possibly bring with them different qualifications and professional experiences that result in decisions that are more likely to be overturned. In this regard, I emphasize quite strongly that the literature is very far from agreement in terms of what constitutes judicial “quality.” At the same time, as Table 1 and some accounts suggest (Lott, 2001; Sen, 2012), African-American lower-court judges are more likely than white judges to be awarded lower qualification ratings from the American Bar Association.
(see Smelcer, Steigerwalt and Jr (2011) for no findings on this point); in addition, a lively scholarly debate has addressed the relative successes of African-American versus white graduates of elite law schools (Sander, 2004; Ho, 2005). Thus, a possibility that must be addressed seriously and delicately is that systematic differences in educational opportunities or professional experiences have translated into some African American judges arriving to the bench with qualitatively different experiences than white judges, and that these differences translate into higher reversal rates. Whether this means that differently qualified judges write opinions reflecting these different experiences (or “qualifications”), or that appeals panels leverage this lack of prestige against judges with non-traditional professional backgrounds is extremely difficult to suss out using this kind of quantitative data. Nonetheless, for the sake of ruling out alternate mechanisms, I address this explanatory hypothesis: *any gap between black and white judges’ reversal rates should attenuate when we compare judges with comparable professional experiences and Bar Association ratings*. Finding such attenuation would suggest a gap driven less by bias, but rather possible differences in professional and educational preparedness.

The last possibility is the possibility of racially motivated implicit bias by appeals panels. To a large extent, a cause for implicit bias using observational data is circumstantial, a stubborn difference that persists despite controls and robustness checks. Nonetheless, as noted, a growing literature suggests that implicit bias against African Americans persists

---

Table 1: Distribution of ABA Qualification Ratings for U.S. District Judges (Johnson to Obama Administrations).

<table>
<thead>
<tr>
<th></th>
<th>Not Qualified</th>
<th>Qualified</th>
<th>Well Qualified</th>
<th>Exceptionally</th>
<th>Well Qualified</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.01</td>
<td>0.43</td>
<td>0.54</td>
<td>0.02</td>
<td>1649</td>
<td></td>
</tr>
<tr>
<td>Whites</td>
<td>0.01</td>
<td>0.41</td>
<td>0.56</td>
<td>0.03</td>
<td>1385</td>
<td></td>
</tr>
<tr>
<td>Blacks</td>
<td>0.01</td>
<td>0.57</td>
<td>0.41</td>
<td>0.00</td>
<td>147</td>
<td></td>
</tr>
<tr>
<td>Hispanics</td>
<td>0.02</td>
<td>0.56</td>
<td>0.41</td>
<td>0.01</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>0.00</td>
<td>0.47</td>
<td>0.52</td>
<td>0.00</td>
<td>279</td>
<td></td>
</tr>
</tbody>
</table>
(and can be measured) within a wide variety of comparable instances, including not just law (Greenwald and Krieger, 2006; Banks, Eberhardt and Ross, 2006; Bagenstost, 2006; Kang, 2004) and politics (Butler and Broockman, 2011), but also employment (Bertrand and Mullainathan, 2004; Fryer and Levitt, 2004a), academia (Ginther et al., 2011), public health (Green et al., 2007), etc. Here, beliefs about blacks’ views could also shape biased views. For example, black judges could simply be perceived to be more liberal (Scherer and Curry, 2010). This in turn could contribute to a biased perception – whether substantiated or not – that black judges decide cases more liberally, thus driving up their reversal rates. Another possibility is that appeals courts view black judges as less qualified simply because of their race; that is, black judges could be perceived to produce opinions of poorer quality, despite no substantive evidence. Ultimately, given how strongly implicit bias has been measured in other areas, and given (despite scholars’ best efforts) continued unexplained gaps between blacks and whites in fields like education (Fryer and Levitt., 2004b) and health care (Jha et al., 2005), this is a possibility that must be considered.

4 U.S. District Judge Data

The data come from the two lower tiers of the federal judiciary – the U.S. District Courts and the U.S. Courts of Appeals. The lower district courts, with 94 courts, hear the largest number of cases. By contrast, the courts of appeals (or “Circuits”) hear cases that parties choose to appeal from the district court. District judges decide cases alone, which makes it easier to determine the impact of a particular judge’s race on appeal and reversal; by contrast, appeals judges nearly always hear cases in panels of three. Also important is that appeals judges have met most of their lower-court counterparts and will therefore be aware of their basic demographics.²

²This assumption is borne out by the fact that higher- and lower-court judges interact personally (by frequently having offices in the same building) and professionally (by participating in judicial conferences and meetings). Dropping the jurisdiction least likely to meet this assumption – the large 9th Circuit – does
Table 2: Racial/ethnic and gender distribution of judicial nominees by President (Johnson through Obama administrations).

<table>
<thead>
<tr>
<th>President</th>
<th>Whites</th>
<th>African Americans</th>
<th>Hispanics</th>
<th>Women</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barack Obama</td>
<td>73</td>
<td>17</td>
<td>11</td>
<td>51</td>
<td>108</td>
</tr>
<tr>
<td>George W. Bush</td>
<td>212</td>
<td>18</td>
<td>27</td>
<td>54</td>
<td>261</td>
</tr>
<tr>
<td>William J. Clinton</td>
<td>229</td>
<td>53</td>
<td>18</td>
<td>88</td>
<td>305</td>
</tr>
<tr>
<td>George H.W. Bush</td>
<td>132</td>
<td>10</td>
<td>6</td>
<td>29</td>
<td>148</td>
</tr>
<tr>
<td>Ronald Reagan</td>
<td>269</td>
<td>6</td>
<td>13</td>
<td>24</td>
<td>290</td>
</tr>
<tr>
<td>Jimmy Carter</td>
<td>151</td>
<td>28</td>
<td>14</td>
<td>29</td>
<td>195</td>
</tr>
<tr>
<td>Gerald Ford</td>
<td>43</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>49</td>
</tr>
<tr>
<td>Richard M. Nixon</td>
<td>170</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>178</td>
</tr>
<tr>
<td>Lyndon B. Johnson</td>
<td>106</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>115</td>
</tr>
<tr>
<td>All</td>
<td>1385</td>
<td>147</td>
<td>95</td>
<td>279</td>
<td>1649</td>
</tr>
</tbody>
</table>

To examine how characteristics of lower-court judges affect case outcomes, I look to data from the Federal Judicial Center (FJC), which makes public key characteristics of all federal judges\(^3\) I coded for the 1,649 judges confirmed from the Johnson through Obama administrations (as of April 2012) each judge’s (1) race or ethnicity, (2) age at investiture, (3) gender, (4) law school attended, and (5) geographic location (Table 2.) I used automated coding to further assess whether each nominee had previously been (6) a former law clerk, (7) a U.S. Attorney or Assistant U.S. Attorney, (8) a Solicitor General or Deputy or Assistant Solicitor General, (9) a state judge (either a state supreme court or state lower court judge), (10) a former federal judge (e.g., magistrate, territorial, or bankruptcy judge), (11) a full-time law professor or law school dean, (12) an attorney in private practice, or (13) a public defender. (Some of these experiences – e.g., legal clerkship – are thought to be high prestige and linked with superior quality; others, e.g., public defenders, are thought the opposite.) I also noted each judge’s American Bar Association qualification rating (historically a four-point scale from “Not Qualified” to “Exceptionally Well Qualified,” with “Exceptionally Well Qualified” dropped in 1989), which could reflect qualitative information not captured not meaningfully alter the results.

\(^3\)http://www.fjc.gov/history/home.nsf/page/judges.html.
<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Whites</th>
<th>Blacks</th>
<th>Black Democrats</th>
<th>Black Republicans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ave Age at Investiture</td>
<td>50.06</td>
<td>50.44</td>
<td>48.55</td>
<td>48.16</td>
<td>49.49</td>
</tr>
<tr>
<td>Female</td>
<td>0.17</td>
<td>0.15</td>
<td>0.27</td>
<td>0.29</td>
<td>0.21</td>
</tr>
<tr>
<td>Nominated by Democrat</td>
<td>0.44</td>
<td>0.40</td>
<td>0.71</td>
<td>1.00</td>
<td>-</td>
</tr>
<tr>
<td>Top 14 Law School</td>
<td>0.30</td>
<td>0.30</td>
<td>0.28</td>
<td>0.31</td>
<td>0.21</td>
</tr>
<tr>
<td>Private Law School</td>
<td>0.52</td>
<td>0.51</td>
<td>0.67</td>
<td>0.69</td>
<td>0.63</td>
</tr>
<tr>
<td>Law Clerk</td>
<td>0.21</td>
<td>0.22</td>
<td>0.14</td>
<td>0.16</td>
<td>0.09</td>
</tr>
<tr>
<td>Law Professor</td>
<td>0.06</td>
<td>0.05</td>
<td>0.12</td>
<td>0.13</td>
<td>0.07</td>
</tr>
<tr>
<td>Private Practice</td>
<td>0.92</td>
<td>0.94</td>
<td>0.76</td>
<td>0.81</td>
<td>0.65</td>
</tr>
<tr>
<td>U.S. Attorney</td>
<td>0.09</td>
<td>0.09</td>
<td>0.03</td>
<td>0.03</td>
<td>0.05</td>
</tr>
<tr>
<td>Assistant U.S. Attorney</td>
<td>0.20</td>
<td>0.19</td>
<td>0.29</td>
<td>0.25</td>
<td>0.40</td>
</tr>
<tr>
<td>Justice Dept Lawyer</td>
<td>0.05</td>
<td>0.05</td>
<td>0.07</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>Public Defender</td>
<td>0.04</td>
<td>0.03</td>
<td>0.10</td>
<td>0.12</td>
<td>0.02</td>
</tr>
<tr>
<td>U.S. Magistrate Judge</td>
<td>0.09</td>
<td>0.08</td>
<td>0.10</td>
<td>0.09</td>
<td>0.12</td>
</tr>
<tr>
<td>U.S. Bankruptcy Judge</td>
<td>0.01</td>
<td>0.01</td>
<td>0.04</td>
<td>0.05</td>
<td>0.02</td>
</tr>
<tr>
<td>State Judge</td>
<td>0.41</td>
<td>0.38</td>
<td>0.55</td>
<td>0.52</td>
<td>0.63</td>
</tr>
<tr>
<td>N</td>
<td>1649</td>
<td>1385</td>
<td>147</td>
<td>104</td>
<td>43</td>
</tr>
</tbody>
</table>

Table 3: Demographics of U.S. District Court nominees named after 1960.

by the quantitative data (Table 1).

For partisanship, I recorded for each judge the identity of the appointing President as well as his or her judicial common space score (Boyd, 2011; Giles, Hettinger and Peppers, 2001; Epstein et al., 2007; Poole, 1998), which relies on the common space score of either the President or of the state’s senior Senator (if of the President’s party). I further coded the law school attended by using the 2001 U.S. News & World Report rankings and dividing them into rank cohorts: (1) elite law schools in the “Top 14,” (2) schools ranked #15-25, (3) #26-50, (3) #51-76, (4) #76-100, and (5) outside of the top 100. This is a rough measure for judges attending law schools in the 1970s and 80s; an assuaging factor is that the top tier’s composition has never changed.

Professional and educational characteristics are only half of the story. To assess the influence of these characteristics on reversal, I also examined case outcomes data via two distinct data sets. First, I used automated coding to collect judge-level reversal statistics
Figure 1: Reversal rates for white and black U.S. District Judges (cases appealed 2000 to July 2012). Dotted vertical lines denote white and black means.

reported by Westlaw, a commercial legal data base, in its “Judicial Reversal Reports.” Included in these reports are (1) the total number of cases, both published and unpublished, for which the district judge wrote an opinion, (2) the total number of cases, published and unpublished, that were appealed for that judge, and (3) how many of these appealed cases, again published and unpublished, were affirmed or reversed. (This is measured as a dichotomous variable, i.e., the case was upheld or it was not.) I used this information to create a new data set that includes for all district judges his or her complete reversal rate from 2000 to July of 2012. This final data set includes reversal rates for 1,053 district judges, of whom 944 are white and 109 are African American. The distribution of judges’ reversal rates is displayed in Figure 1.

Because the identity of the judges hearing the appeal could influence the case outcome, and because the Judicial Motion Reports by their aggregate nature do not contain this information, I examine an extant data set of appealed cases collected by Songer, Kuersten
and Haire (2007). These data include 10,957 randomly selected published appeals cases decided between 1996 and 2002,\textsuperscript{4} of which 537 were originally decided by black district judges and 10,420 by white judges. I also have data on the three judges hearing the appeal, its substantive legal issue area (e.g., “Civil Rights,” “Criminal Law,” “Economic Activity,” “Labor”), and whether the lower-court opinion was (1) upheld or (2) reversed.

Here, a fact helpful to identification is that incoming cases in (1) district courts and in (2) the appeals stage are assigned to judges (or panels) on a fairly random basis. Although the randomization can be informal, this longstanding practice makes it impermissible for federal judges to request to hear particular kinds of cases. Thus, conditional on jurisdiction, (1) cases heard by black lower-court judges should on average be similar to those heard by white judges (i.e., there should be balance in case characteristics between cases heard by black judges versus those heard by white judges) and (2) appeals panels hearing cases written by black judges should on average be similar to appeals panels hearing cases written by white judges (that is, potentially biased judges cannot request to hear cases decided by lower-court black judges). I present results that suggest that the randomization is working below. I also at times control for issue area in the case-level data; the substantive results are unaffected.

5 Race as a Predictor of Reversals

I now turn to the key question: whether black judges are overturned more or less than their white colleagues. I do this analysis twice, once looking at the new data on judges’ overall reversal rates and again looking at case-level data from Songer, Kuersten and Haire (2007). When examining the judges’ reversal rates, which include all published and unpublished cases appealed 2000-July 2012, an important consideration is that the number of cases a judge hears varies by jurisdiction and length of service. For example, a judge retiring in 2001 will

\textsuperscript{4} Although perfect overlap with the Westlaw data is preferable, the Songer, Kuersten and Haire (2007) data only go through 2002.
have fewer cases included versus a judge serving the entirety of 2000-2012. An ordinary least squares specification with the reversal rate as the outcome would therefore violate basic OLS assumptions: the variance of the outcome would clearly vary according to whether the judge had 1 case appealed or 180. For the reversal data, I therefore take a weighted least squares (WLS) approach by weighing each judge by the square root of the number of cases he or she had appealed (Lewis and Linzer, 2005). Different weighting schemes results in similar inferences; I also explore the issue of race-based differences in appeals rates below.

For the analyses looking at the case-level Songer, Kuersten and Haire (2007) data, I use a logit specification, with the outcome variable being whether a case was reversed (1) or upheld (0). Here, I also include judge-specific random effects to account for the fact that one judge might hear multiple cases (and observations are not independent). In both, to guard against the possibility that the results could be model dependent, I fit a variety of models with a variety of independent variables (here and in the following sections), including dummies for appointing President, the District Court where the judge sits, the Circuit hearing the appeal, and several demographic characteristics.

Tables 4 (judge-level reversal rates) and 5 (case-by-case reversals) present these results. Focusing on Table 4, the effect for black district judge is always positive and statistically significant, ranging in magnitude from 2% to 3%. Substantively this means that black judges have a reversal rate that is between 2% and 3% higher than whites. This difference persists after taking into account structural characteristics that could explain discrepancies in reversal rates, such as black and white judges being appointed by different Presidents (Table 4 Models 2–6), living in different jurisdictions (Models 3–6), being of different ages on confirmation (Models 4 and 6), and having in their ranks different proportions of men and women (Model 4 and, showing that an interaction between race and gender is insignificant, in Model 6). I also test whether the effect increases significantly when the appeals court is located in the South, by including dummy variables for the Richmond-based 4th Circuit,
Table 4: Weighted OLS regression results, U.S. District Court judges’ reversal rates (continuous between 0 and 1) for cases decided between 2000-2012.

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
<th>Model 7</th>
<th>Model 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.31*</td>
<td>0.32*</td>
<td>0.47*</td>
<td>0.48*</td>
<td>0.40*</td>
<td>0.40*</td>
<td>0.40*</td>
<td>0.40*</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Black District Judge</td>
<td>0.03*</td>
<td>0.03*</td>
<td>0.02*</td>
<td>0.02*</td>
<td>0.03*</td>
<td>0.03*</td>
<td>0.03*</td>
<td>0.02*</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Female</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Age</td>
<td>0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Female × Black District Judge</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>South</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>South × Black District Judge</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Circuit Dummies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>President Dummies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>District Court Dummies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>N</td>
<td>1053</td>
<td>1053</td>
<td>1053</td>
<td>1053</td>
<td>1053</td>
<td>1053</td>
<td>1053</td>
<td>1053</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.01</td>
<td>0.02</td>
<td>0.46</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.01</td>
<td>0.01</td>
<td>0.45</td>
<td>0.48</td>
<td>0.48</td>
<td>0.48</td>
<td>0.48</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

* indicates significance at $p < 0.05$

The key results are substantiated by similar findings based on the case-level data, which are presented in Table 5. Again, cases written by black lower-court judges are more likely to be reversed on appeal, and the effect is robust to the inclusion of district-level dummies and dummy variables for the appointing President.\(^5\) In terms of predicted probabilities, Table 5 actually suggests even stronger effects: a case written by a black judge on average may have up to a 5% to 7% greater chance of being reversed than a case written by a white judge. These are substantively significant differences: considering that the average district judge has approximately 190 cases appealed in a 12-year period, this racial gap results in a difference between black and white judges of approximately 865 to 2,019 cases. That is, up

---

\(^5\)The results are also robust to the inclusion of age and gender and, in results not shown here, to the inclusion of the number of years on the bench (experience) the judge had before hearing the case in question.
Table 5: Logit regression regression results, case-by-case basis. Outcome variable is whether case reversed (1) or upheld (0). Data are randomly selected subset of published cases (Songer, Kuersten and Haire, 2007), 1996-2002. Judge random effects included.

to 2,000 black-authored cases would still be on the books if black judges were overturned at whites’ lower rates.

Two points are worth further mention. First, I note that the substantive interpretations do not change depending on whether we include District Court dummies (Table 4, Models 4–6) or Circuit dummies (Table 4, Model 3). However, the explanatory power from the model increases swiftly when we include any kind of control for jurisdiction. (Table 4’s overall $R^2$ moves from around 0.02 to around 0.53, Models 3–6.) I move forward in the remaining analyses by including dummy variables for district court. This has the additional benefits of (1) controlling for fluctuations in case dockets among districts (e.g., the Southern District of California versus the District for Alaska), and also of (2) allowing the assumption that cases are randomly assigned to be made more safely (as cases are usually assigned randomly, but only within district). Second, the substantive results are similar regardless of whether we look at overall reversal rates (Table 4) or randomly selected case-level data (on published cases, Table 5). In the analyses that follow, I primarily use the overall reversal rates, using
the case-level data to analyze instances when the composition of the appeals panel is thought to play a salient role. I do so because the overall reversal rate data is both more conservative and, more importantly, because it includes the universe of appealed cases, not just a random subset of published cases.\textsuperscript{6}

These results provide support for the first hypothesis that black judges will be more likely to be overturned on appeal. I turn now to exploring possible reasons behind this difference, beginning with the theory that appeals courts are actually hearing cases that vary in number and type according to the race of the lower-court judge.

6 Mechanism 1: Differences in Cases Appealed

The first explanation I consider is that cases authored by black lower-court judges will be more likely to be appealed, perhaps as a result of differences in black-white voting, or from an implicit bias against black judges by legal practitioners (as suggested by the state-courts literature (Gill, Lazos and Waters, 2010)). Although I control for the number of cases appealed (via weighting in Table 4), it is possible that a persistent difference in both the number and the type of cases appealed could introduce bias. That is, having more cases appealed from black judges may mean that more of them will be reversed; likewise, if more civil rights cases are appealed from black judges, and if civil rights cases are more likely to be reversed, this would create the impression of race-based reversal when none in fact exists.

I first examine the number of cases appealed, and whether they vary by race of lower-court judge. For these analyses, the outcome variable is the number of cases each district judge had appealed 2000-2012. To take into account the fact that some judges produce more or fewer authored opinions, I also include the number of opinions each judge produced during this time period as a control variable (Table 6, Models 2–4). Further, to take into account

\textsuperscript{6}Note that including unpublished cases has the effect of reducing possible bias stemming from the decision to publish, but it also likely has the effect of presenting a more conservative overall estimate of the black judges’ effect, as unpublished cases have been shown to display less variance (Keele et al., 2009).
Table 6: OLS regression results. Outcome variable is total number of cases appealed from each judge (between 2000 and 2012).

Surprisingly, Table 6 demonstrates that black judges are no more likely to have their opinions appealed than are white judges: under all model specifications, the substantive difference between black and whites is negligible and never statistically significant. There are two further items of note. First, as would be expected, the total number of cases each judge produces is itself (slightly) predictive of the number of cases appealed – that is, judges who write more opinions have more of them appealed. Second, what explanatory power we have in the model comes not from the race of the lower-court judge (nor his or her party, Models 1, 2, and 4), but from the addition of dummies for the district court (Models 3 and 4). Thus, to the extent that appeals rates vary, they do so by jurisdiction and not by the race of the lower court judge. Along with controlling for the number of cases appealed, this assuages the concern that the results are simply driven by more cases being appealed from
Table 7: $\chi^2$ tests of difference between black and white judges’ cases across legal issue area.

<table>
<thead>
<tr>
<th>Circuit</th>
<th>$p$-value</th>
<th>Significant?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>0.73</td>
<td>No</td>
</tr>
<tr>
<td>2d</td>
<td>0.71</td>
<td>No</td>
</tr>
<tr>
<td>3d</td>
<td>0.54</td>
<td>No</td>
</tr>
<tr>
<td>4th</td>
<td>0.33</td>
<td>No</td>
</tr>
<tr>
<td>5th</td>
<td>0.11</td>
<td>No</td>
</tr>
<tr>
<td>6th</td>
<td>0.22</td>
<td>No</td>
</tr>
<tr>
<td>7th</td>
<td>0.43</td>
<td>No</td>
</tr>
<tr>
<td>8th</td>
<td>0.17</td>
<td>No</td>
</tr>
<tr>
<td>9th</td>
<td>0.51</td>
<td>No</td>
</tr>
<tr>
<td>10th</td>
<td>0.11</td>
<td>No</td>
</tr>
<tr>
<td>11th</td>
<td>0.36</td>
<td>No</td>
</tr>
<tr>
<td>DC</td>
<td>0.17</td>
<td>No</td>
</tr>
</tbody>
</table>

black lower-court judges.

However, the other concern is perhaps that the number of cases might not vary, but the kind of case does. Here, the particular concern that black judges’ rulings on racially salient judges may be more (or less) likely to be appealed. To address this issue, I supplement Table 6 by using data from Songer, Kuersten and Haire (2007) to analyze whether case dockets involving published appeals vary according to the race of the lower court judge. I use a series of simple $\chi^2$-square tests to examine a null hypothesis of no relationship between the race of the lower-court judge and substantive case legal issue, conditional on jurisdiction (appeals court). The categories I test are the seven issue areas coded by Songer, Kuersten and Haire (2007): criminal, civil rights, First Amendment, due process, privacy, labor relations, economic activity, and miscellaneous (and also a “not ascertained” category). Across all of the jurisdictions, I cannot reject the null hypothesis that there is no relationship between lower-court judge race and type of case heard on appeal. Taken together, this leads me to reject the second explanation. There is no support that attorneys either (1) appeal from opinions written by black judges at higher rates, or (2) choose to appeal different kinds of cases depending on the race of the lower-court judge. Thus, to the extent that black judges
are more likely to be reversed on appeal, it is unlikely to be due to the number and nature of the kinds of cases being appealed.

7 Mechanism 2: Possible Differences in “Quality”

I turn next to another possible explanation. The FJC data show that black and white judges differ on average in terms of some previous employment, qualification scores, and law schools attended (Tables 1 and 3). For example, fewer black judges have private practice experience compared to white judges (76% compared to 94%), while more have state judge experience (55% compared to 38%). Black judges are also less likely to receive higher qualification ratings from the American Bar Association; 41% of them receive a high “Well Qualified” rating compared to 56% of white judges. (Discrepancies in qualification ratings are explored in Lott (2001) and Sen (2012), with Smelcer, Steigerwalt and Jr (2011) providing evidence to the contrary.) However, there is disagreement in the literature on whether objective criteria can ever determine what makes a “good” judge, whether such criteria are useful predictors of reversal, and whether some factors (e.g., public defender experience) could cut both ways. Nonetheless, addressing these issues is essential to possibly understanding why reversal rates for black and white judges differ, and I do so by including both objective measures of experience and education, and also by including subjective measures such as ABA ratings.

As in Table 4, the outcome variable for this analysis is the judge’s reversal rate from 2000-2012, across all cases (published and unpublished). I again include dummy variables for the (1) identity of the appointing President and (2) district court. The results are displayed in Table 8. In Model 1, I include the ABA ratings as dummy variables. In Model 2, I include various educational variables – including rank cohort of law school attended. (Here and in subsequent models the excluded category is attending a Top 14 (“T14”) law school, considered the elite group; a large number of judges attended one of these schools,
particularly Harvard (121 judges) and Yale Law Schools (54 judges).) Model 3 includes professional experience, such as whether the judge had been in private practice or was a law clerk. And, finally, Model 4 includes whether the judge had served in a judicial capacity before, perhaps as a federal magistrate or state judge (either state supreme or state lower court judge).

Despite the inclusion of both the subjective and objective criteria, the inferences do not change: black district judges are still more likely to be overturned on appeal than are white judges – with an increase in their reversal rate of approximately 2%. Perhaps surprising is the fact that relatively few of the educational measures and professional experiences predict reversal: there are few differences between those who went to high-ranked law schools and those who didn’t, those who were US attorneys and those who weren’t, etc. The only exception, perhaps, are former state judges (who are more likely to be reversed) and former law clerks (slightly less). The models are by observational standards fairly predictive, with $R^2$ values close to 0.50. However (as before), most of the explanatory power comes not from the professional or educational variables, but rather from the addition of dummy variables for appointing President and for jurisdiction.

Although the results do not attenuate with the inclusion of these professional and educational characteristics, we must consider whether other unmeasurable or qualitative traits are driving the results. Here, some literature suggests that even black graduates of high-performing law schools do not perform at the level of their white peers (Sander (2004); but see Ho (2005) for a rebuttal). For this analysis, some traits that are not captured in the FJC data are LSAT score, law review membership, law school rank, Order of the Coif, bar passage, or writing ability. I do not address this debate on black achievement specifically, but I cite three reasons why this might be less of a concern with this analysis. The first is that few of the prestige-oriented variables that are included in the federal data are actually predictive at all of reversal rates: being a U.S. Attorney, graduating from a top law
<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.44*</td>
<td>0.47*</td>
<td>0.46*</td>
<td>0.46*</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Black District Judge</td>
<td>0.02*</td>
<td>0.02*</td>
<td>0.02*</td>
<td>0.02*</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Female District Judge</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Age of District Judge</td>
<td>0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
<td>0.00*</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Law School Ranked #15-25</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Law School Ranked #26-50</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Law School Ranked #51-75</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Law School Ranked #76-100</td>
<td>-0.01</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Law School Ranked #101+</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Law Clerk</td>
<td>-0.01*</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Law Professor</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Private Practice</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>US Attorney</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Assistant US Attorney</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Justice Department Experience</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Public Defender</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Federal Magistrate</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Federal Bankruptcy</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>State Judge</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>ABA Ratings Dummies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>President Dummies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>District Court Dummies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>N</td>
<td>1053</td>
<td>1053</td>
<td>1053</td>
<td>1053</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.54</td>
<td>0.54</td>
<td>0.55</td>
<td>0.55</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.48</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
</tr>
<tr>
<td>Resid. sd</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>

* indicates significance at $p < 0.05$

Table 8: Weighted OLS regression results, U.S. District Judges’ reversal rates (continuous between 0 and 1) for cases decided between 2000-2012.

school, or even having served as a full-time law professor or dean do little to predict judges’ reversal rates. Being a law clerk is only fleetingly so, and its effect is quite minuscule compared to the variance explained by simply conditioning on jurisdiction (or even compared to the effect of a lower-court judge being black). An appropriate supposition is that includ-
ing similar prestige-oriented variables (e.g., law review membership) would result in similar non-significance.

Second, although I do not have access to data like class rank or bar passage (which are not made public by the FJC), I do condition on attributes clearly predicated on those marks of success: few advance to legal clerkships, law professorships, or U.S. Attorney positions without having achieved some combination of high class rank (GPA), law review membership, and bar passage. Presumably controlling for these professional experiences also controls to some extent for these unrecorded traits. Lastly, I also condition on ABA ratings – which purport specifically to assess a judicial nominee’s “professional competence,” a qualitative assessment based on quality of legal reasoning, class rank, law review membership, and bar passage. Thus, these are characteristics that would likely be reflected via a candidate’s ABA rating, and the results are robust to the inclusion of this variable. Taken together, this leads me to reject this possible explanation. I do so more tentatively than the first, however, because of the possibility of omitted variables.

8 Mechanism 3: More Liberal Voting

The results suggest that the racial gap persists despite controlling for differences in backgrounds – including both objective experiences and subjective qualification ratings. Another plausible explanation is that black judges are more liberal across the board than are comparable white judges. For example, some literature (Asmussen, 2011) suggests that Presidents who appoint minorities take the opportunity to appoint more ideologically driven individuals than they would otherwise. For African Americans, this would bring to the bench more left-leaning (or right-leaning in the case of Republican presidents) black candidates, who would then be overruled more by moderate appeals panels across all kinds of legal issue areas. I note some evidence of this in Figure 2, which shows the judicial common space scores of black and white judges by party (Boyd, 2011; Giles, Hettinger and Peppers, 2001; Epstein
Figure 2: Judicial common space scores of black and white district judges, by race and party

et al., 2007; Poole, 1998). Black judges have more left-leaning judicial common space scores, which in turn raises the possibility that they write opinions that are more liberal (and hence, reversed at higher rates by more centrists appeals panels).

Testing district judge ideology directly. To analyze further the role of lower-court ideology, Table 9 explores reversal rates with four model specifications: (1) district judge party included, (2) district judge party and race interacted, separately for (3) district judges appointed by a Democrat, and for (4) district judges appointed by a Republican. (A separate set of analyses including the judges’ judicial common space score in the model, not shown, yielded identical results – not surprising, as conditioning on President and jurisdiction has the effect of usually conditioning on judicial common space score.) To be clear, that there are no differences between Democrats and Republicans cannot be ruled out (Model 2); an
interaction between race of district judge and party of appointing President is not significant under any model specifications. However, what we can safely say is that the effect appears driven by Democratic appointed judges. Among Democratic appointed judges, the reversal rate increases by approximately 3% for black judges compared to whites, and this difference is statistically significant; the effect is robust to the inclusion for the variables associated with prior experience and also to the inclusion of dummy variables for appointing President and jurisdiction (district court). Among Republican judges, however, the effect is close to 0% and never significant. Neither is the gap driven by systematic differences between black Democrats and Republicans, for example in terms of their previous professional experience: the models all include controls for professional experience and ABA ratings and, as Table 3 demonstrates, black Democrats and black Republicans do not differ in ways that would sug-
<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black District Judge</td>
<td>0.39*</td>
<td>0.39*</td>
<td>0.38*</td>
<td>0.35*</td>
<td>0.28°</td>
<td>0.37°</td>
</tr>
<tr>
<td></td>
<td>(0.15)</td>
<td>(0.15)</td>
<td>(0.15)</td>
<td>(0.15)</td>
<td>(0.16)</td>
<td>(0.22)</td>
</tr>
<tr>
<td>Panel Median Ideology</td>
<td>−0.42*</td>
<td>−0.42*</td>
<td>−0.40*</td>
<td>−0.39*</td>
<td>−0.38*</td>
<td></td>
</tr>
<tr>
<td>(Conservative)</td>
<td>(0.13)</td>
<td>(0.13)</td>
<td>(0.14)</td>
<td>(0.14)</td>
<td>(0.14)</td>
<td></td>
</tr>
<tr>
<td>Panel Ideology × Black</td>
<td>−0.39</td>
<td>−0.39</td>
<td>−0.36</td>
<td>−0.39</td>
<td>−0.35</td>
<td></td>
</tr>
<tr>
<td>District Judge</td>
<td>(0.33)</td>
<td>(0.41)</td>
<td>(0.41)</td>
<td>(0.41)</td>
<td>(0.41)</td>
<td></td>
</tr>
<tr>
<td>Republican District</td>
<td>−0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Republican Majority</td>
<td>−0.17*</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Republican Panel × Black</td>
<td>−0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District Judge</td>
<td>(0.26)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.013</td>
<td>0.014*</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female District Judge</td>
<td>0.19</td>
<td>0.16</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.15)</td>
<td>(0.16)</td>
<td>(0.15)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge-Specific Random</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District Court Dummies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>President Dummies</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>ABA Ratings Dummies</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educational Controls</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional Controls</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>3866</td>
<td>3866</td>
<td>3866</td>
<td>3866</td>
<td>4631</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standard errors in parentheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>* indicates significance at $p &lt; 0.05$</td>
</tr>
<tr>
<td>° indicates significance at $p &lt; 0.10$</td>
</tr>
</tbody>
</table>

Table 10: Logit regression regression results, case-by-case basis. Outcome variable is whether case reversed (1) or upheld (0). Data are randomly selected subset of published cases (Songer, Kuersten and Haire, 2007), 1996-2002. Judge random effects included.

gest that black Republicans are somehow stronger or more experienced judicial candidates. (If anything, a higher share of black Democrats attended elite “T14” law schools and served as law clerks.)

**Testing relationship to appeals panel ideology.** These results are consistent with black judges simply being more liberal than other judges: compared to white Democrats, African American judges are more liberal and therefore more likely to be reversed. (The results are actually not consistent with a parallel story, one in which black Democrats are
more liberal, and black Republicans more conservative – i.e., minority candidates are more “extreme.” If this story was true, then we would see the effect regardless of party, and not for Democrats only.) To further evaluate the possibility that black judges being more liberal is driving the result, I turn to the case-level data from Songer, Kuersten and Haire (2007) analyzed in Table 5. If the effect is driven by black judges being more liberal than whites, then we would expect that (1) not only would black Democrats be more likely than white Democrats to be reversed (as we see in Table 8), but also that (2) the difference between black and white judges in terms of reversal rates increases when black judges’ opinions are reviewed by conservative 3-judge appeals panels. That is, the gap between black and white judges should grow with more conservative reviewing courts.

I check this in the case-level data by analyzing (1) an interaction between the lower-court judge’s race and the median judicial common space score of the presiding appeals panel, and (2) an interaction between the lower-court judge’s race and whether Republicans are a majority on the appeals panel. If black judges being more liberal is what drives their increased reversal, then we would expect a positive relationship in the interaction terms. However, as Table 10 demonstrates (using several model specifications), the interaction of the black judges’ variable and the median ideology and partisanship of the appeals panel is never significant; that is, the “black judges” effect does not vary according to the ideology or partisanship of the appeals panel. If anything, the negative coefficient suggests that the effect decreases with when the 3-judge panel becomes more conservative. The explanation behind this is unclear, and the interaction is never significant. However, it does provide additional evidence against the theory that the fact that black judges’ increased reversal is being driven by these black judges being more liberal across the board.
9 Mechanism 4: Voting on Certain Issues

A remaining possibility is that black judges vote differently than white judges, but that they do so only with regard to cases having a significant racial, ethnic, or civil rights dimension. This could include substantive issue areas involving affirmative action and civil rights (Kastellec, 2012) or, possibly criminal cases involving race-based defenses by black defendants. We may expect black judges to be overruled most frequently in these areas, where their views might differ the most from whites. Or, we may expect that black judges may be upheld more in these areas, with appeals judges being deferential to black judges on racially sensitive cases. In either scenario, we would expect that the “black judges effect” would vary significantly between possibly racially-salient areas (e.g., civil rights law) and others.

To test this possibility, I include in my analyses of the case-level data a dummy variable for civil rights-related cases, as coded by Songer, Kuersten and Haire (2007). (In models not shown, I also control for all seven of Songer et al’s issue areas; the inferences are not affected.) In several of the models, I interact the case issue area with the race of the lower court judge. For the “issue area” explanation to hold sway, the racial gap should differ across the areas identified by the judicial politics literature as being particularly racially salient; that is, we would see a significant relationship in the interaction of black lower-court authorship and civil rights issues. The results show, however, that the effect of black authorship on a case’s probability of being upheld actually varies little by issue area: there is no difference across civil rights cases and non-civil rights cases. Thus, the data provide no evidence for the proposition that black judges are being overturned at greater or lesser rates (compared to white judges) within different legal categorizations. The racial gap is therefore unlikely to be driven exclusively by black judges deciding racially salient cases differently.
<table>
<thead>
<tr>
<th>Model</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black District Judge</td>
<td>0.22*</td>
<td>0.22*</td>
<td>0.23*</td>
<td>0.34*</td>
<td>0.33*</td>
<td>0.37*</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.11)</td>
<td>(0.12)</td>
<td>(0.14)</td>
<td>(0.15)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Civil Rights Case</td>
<td>0.14°</td>
<td>0.15°</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.08)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Civil Rights Case × Black District Judge</td>
<td>−0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.27)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Black on Panel</td>
<td>0.09</td>
<td>0.072</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Blacks on Panel</td>
<td>0.60</td>
<td>0.91*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.40)</td>
<td>(0.44)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Black on Panel × Black District Judge</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.31)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Blacks on Panel × Black District Judge</td>
<td>−1.81†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Majority Black Panel</td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.44)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Majority Black Panel × Black District Judge</td>
<td>−1.85†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Issue Area Dummies: ✓
Judge-Specific Random Effects: ✓ ✓ ✓ ✓ ✓ ✓
President Dummies: ✓ ✓ ✓ ✓ ✓ ✓
District Court Dummies: ✓ ✓ ✓ ✓ ✓ ✓
ABA Ratings Dummies: ✓ ✓ ✓ ✓ ✓ ✓
N: 7295 7295 7295 4635 4635 4635

Standard errors in parentheses
* indicates significance at $p < 0.05$
° indicates significance at $p < 0.10$
† indicates significance at $p < 0.15$

Table 11: Logit regression regression results, case-by-case basis. Outcome variable is whether case reversed (1) or upheld (0). Data are randomly selected subset of published cases (Songer, Kuersten and Haire, 2007), 1996-2002. Judge random effects included.

10  Mechanism 5: Implicit Bias by Appeals Panels

An explanation that must be considered is that appeals panels somehow implicitly rely on the race of the lower-court judge in reaching decisions. On the one hand, this explanation has the deepest and most troubling normative implication, and challenges the fairness and race neutrality of the judiciary. On the other, such a finding would perhaps be unsurprising, as studies have teased out implicit biases against racial minorities in prominent economic,
Here, I consider one possible manifestation – whether having a black presence on an appeals panel attenuates the effect. This could happen as white judges become more sensitive to any possible discriminatory tendencies, or as black judges raise possible concerns about bias. (On this point, Kastellec (2012) provides evidence that having a black judge on an appeals panel will change the way that the panel votes on affirmative action issues.) To test this theory, I evaluated how the racial gap varies across different racial appeals panel compositions – zero, one, or two black judges on the three-judge higher-court panels. As before, I use mixed-effect logit model, with an interaction between race of the lower-court judge and the number of black circuit judges hearing the appeal. No cases in the data were heard by an all-black 3-judge panel.

Results from this analysis are presented in Table 11, Models 3–6. Because of the exceedingly low numbers of black judges on appeals courts, and because of the fact that these judges very rarely sit together, it is impossible to distinguish statistically panels heard by zero, one, or two black judges on the appeals. Nonetheless, the trend is obvious – having 0 or 1 black judge is associated with a constant size of the “black judges effect.” However, when increasing the number of black judges on the appeals panel to two, where black judges represent the majority, the effect disappears and not only loses significance, but also moves in a more positive direction. That is, black lower-court judges appear increasingly less likely to be overturned by appeals panels where blacks are the majority. To be clear, the exceedingly small number of cases where blacks constitute a majority means that this effect is not statistically significant at the 5% level (we have a p-value of 0.12). I also note that this doesn’t rule out the possibility that black appeals judges are more liberal than white judges, an implication that would also explain the results on more ideological grounds (although ad-

---

7 Of the cases decided between 1960 and 2002 that were coded by the Songer, Kuersten and Haire (2007) team, 88% had no black judges on the 3-judge appeals panel, 11% had one black judge, and just under 1% had two.
ditional controls for the median panel ideology do not change the results obtained in Models 3–6).

11 Conclusion

The results show that discrepancies exist in how appeals courts review cases, with black judges consistently being more likely to be reversed than whites. This racial gap is robust and persists once we control for possible proxies for judicial “qualifications” – e.g., quality of legal education, age, professional experience, and American Bar Association ratings. Controls for the partisanship of the lower-court judge, as well as for the partisanship and racial composition of the reviewing appeals panel do not affect the results. The discrepancy in reversal rates between black and white judge is consistent across issue area and across jurisdictions. And gap is significant: up to 2,000 federal court cases would still be on the books if black judges were overturned at whites’ lower rates.

The reasons behind this persistent difference are not straightforward. Although having blacks on the reviewing panel appears to attenuate the effect, there are too few black appeals court judges to make meaningful inferences. The difference appears not to be driven by black judges voting differently on certain cases. At best, we have suggestive evidence that black judges are more liberal than otherwise similar whites, but the fact that black judges are overturned at roughly equal rates by conservative and liberal appeals panels casts doubt on this being the sole explanation. A more likely explanation is that this discrepancy is driven by a variety of factors – perhaps due to black judges being more liberal in ways unmeasured by extant ideology measures (accompanied or perhaps buttressed by the perception of black judges being more liberal (Scherer and Curry, 2010)), but also perhaps due to implicit biases by higher courts (perhaps based on the perception that black judges are less qualified).

A point worth emphasizing is that the gap between black and white judges attenuates at times, but never fully disappears. To this extent, this is a situation somewhat similar to
stubborn racial gaps in other important areas, such as education (Fryer and Levitt., 2004b) and health care (Jha et al., 2005), on which scholars have devoted substantial time and attention. The implications for this particular gap are striking. Since John F. Kennedy, American Presidents have actively sought to appoint judges of color – not just African Americans, but also Hispanics, and Asian Americans – to the nation’s highest courts. At the state and international level, too, efforts are underway to increase the proportion of judges from under-represented communities. The racial gap demonstrated here, however, calls into question whether the mere appointment of these individuals is enough, and whether more ought to be done to ensure equality in the nation’s courts. After all, if certain judges are being systematically overturned more often, then this raises questions about their long-term impact on the law, legal precedent, and the legal system.

The results presented in this paper actually represent the tip of the iceberg in exploring the components of judicial evaluation and its relationship to descriptive representation – a topic previous unexplored in the judicial politics literature. I touched upon just one singular ascriptive characteristic: the race (black or not) of lower court judges. Whether a judge is African American is, however, just one facet of judicial identity, and we may think that the similar effects may exist for other racial or ethnic groups (Asian Americans, Hispanics), religious groups (Jews, Catholics), and genders – not to mention multiple combinations of these identities. In addition, if we think that heuristics or personal familiarity may play a role in how appeals panels reach decisions, then maybe we would find different rates of overturning between judges who attended the same law school or are otherwise knowledgeable or friendly – that is, that a personal connection strengthens a bond that makes reversal less likely. Further research should help clarify the extent to which these and other attributes might play a role in appellate review.

In addition, this is a study that relies on a quantitative analysis of aggregated data. Still remaining is a closer, qualitative look at the opinions authored by both black and white
lower-court and appeals judges. Do black judges use different legal reasoning or articulate legal principles in a different way? Do black judges rely on particular arguments in defining their opinions? Does the language used by appeals panels differ according to the identity of the legal actors involved? The analyses presented here suggest that there could be something qualitatively different about those opinions written by black judges, as well as some qualitative differences in how appeals panels review cases decided by black judges. Given the results of this analysis, a qualitative examination into these issues would further shed light into why black judges are more likely to be overruled, and why this racial gap is so persistent.

References


Asmussen, Nicole. 2011. “Nontraditional Judicial Nominees: President’s Delight and Senators’ Dismay?” Legislative Studies Quarterly. 23


Gill, Rebecca, Sylvia R. Lazos and Mallory Waters. 2010. “Sacrificing Diversity for ‘Quality’: How Judicial Performance Evaluations are Failing Women & Minorities.”. 5, 6, 17


34


Krislov, Samuel. 1974. Representative Bureaucracy. Prentice-Hall. 3


Savage, Charles. 2011. “Ratings Shrink President’s List for Judgeships.”. 4


Segal, J.A. and H.J. Spaeth. 2002. The Supreme Court and the attitudinal model revisited. Cambridge Univ Pr. 3


Songer, Donald, Ahslyn K. Kuersten and Susan B. Haire. 2007. “The United States Courts of Appeals Database.”. 2, 12, 13, 14, 16, 19, 26, 27, 28, 29, 30


