Trading Diversity? Judicial Diversity and Case Outcomes in Federal Courts*

Ryan Copus[†] Ryan Hübert[‡] Paige Pellaton[§]

May 30, 2024

Abstract

Are federal lawsuits resolved differently based on the race or gender of the judges assigned to hear them? Recent empirical research posits that women and judges of color decide cases more liberally, at least in some identity-salient areas of law. However, these studies analyze small numbers of cases and judges, and use research designs that limit their causal interpretations. Using an original dataset of all civil rights cases filed in 20 federal district courts over multiple decades and a strong causal identification strategy, we find that assignment of cases to judges of color or women has no statistically significant effect on case outcomes among Democratic appointees. However, it causes more conservative outcomes among Republican appointees. We explain these results with a theory of bargaining over judicial appointments in which Republican presidents take advantage of Democrats' preference for diversity on the bench to appoint more conservative judges.

Keywords: judicial politics, trial courts, causal inference, diversity

Replication files are available in the Harvard Dataverse at:
https://doi.org/10.7910/DVN/I8B3VS.

10,579 words

*Authors are listed in alphabetical order. We thank Sarah Anzia, Rachel Bernhard, Chris Cotropia, Scott Daniel, Sean Gailmard, Luzmarina Garcia, Mark Hurwitz, Jonathan King, Gabe Lenz, Adam Pah, David Schwartz, Moses Shayo, Rorie Solberg, Jörg Spenkuch, Laura Stoker, Dvir Yogev, and audience members at UC Berkeley, University of Chicago, LSE, Oxford, MPSA 2022, APSA 2022, and CELS 2022 for their helpful feedback. We are especially grateful to the Systematic Content Analysis of Litigation EventS (SCALES) Open Knowledge Network (https://scales-okn.org/) for the use of their data. All errors and omissions are our own.

[†]Ryan Copus is an Associate Professor at the University of Missouri–Kansas City School of Law, 500 E. 52nd Street, Kansas City, MO 64110. Email: copusr@umkc.edu.

[‡]Ryan Hübert is an Assistant Professor of Political Science at the University of California, Davis, One Shields Avenue, Davis, CA 95616. Email: rhubert@ucdavis.edu. **Corresponding author.**

[§]Paige Pellaton is a Ph.D. candidate in Political Science at the University of California, Davis, One Shields Avenue, Davis, CA 95616. Email: ppellaton@ucdavis.edu.

There will not be an ideological blood test, like there was during the Reagan and Bush years, to see if the candidate is a moderate or liberal. But there will be an insistence upon diversity.

– Senate Judiciary Committee Chair Joe Biden in 1993[1]

The beginning of the Biden Administration was marked by a concerted effort to diversify the federal judiciary. In addition to promising to nominate a Black woman to the Supreme Court—and doing so—the newly-elected president set to work nominating a historically diverse slate of judges to the lower courts. Two-thirds of the judges appointed during his first year in office were judges of color and 80 percent were women, far surpassing the efforts of his predecessors to increase diversity on the bench in their first years. Biden's insistence on a more inclusive judiciary, following President Trump's disinterest in the issue, has reignited a debate about the importance of having more women and judges of color on the bench.

There are many claims made by observers, activists, and political actors about the impacts of diversifying the federal bench. Of particular interest is the extent to which judges from historically underrepresented racial, ethnic, or gender backgrounds approach cases differently than the White men who have traditionally been appointed to the bench. A large body of quantitative research has developed to examine this question. Notwithstanding a handful of exceptions, the recent conventional wisdom that has emerged from this research is that women and judges of color resolve cases in a more liberal manner, at least in some identity salient areas of law (see ch. 3 of Friedman et al. 2020).

Our analysis of an original dataset of federal civil rights cases makes three important contributions. First, we focus attention on the federal district courts, which have received disproportionately less attention from scholars than federal appeals courts. A large majority of federal judges are seated in district courts, including the vast majority of women and judges of color. The district courts are also the workhorses of the federal judiciary, providing final resolution in more than 90 percent of federal lawsuits. As U.S. District Judge Henry N. Graven once famously said, "The

1. Quoted in Labaton (1993).

people of this district either get justice here with me or they don't get it at all" (quoted on p. 1 of Carp and Rowland 1996).

Second, we take great effort to eliminate post-treatment bias. Many prior research findings are vulnerable to post-treatment bias because they are based on datasets that exclude certain cases based on their outcomes (e.g., excluding settled cases, see Hübert and Copus 2022). To be fair, large and comprehensive datasets of federal cases are difficult to create since important data on federal cases are contained in court records that are kept behind an expensive paywall (called "PACER", see https://pacer.uscourts.gov). Our data collection process underscores the seriousness of this barrier: long waits for each court to decide whether to issue a researcher fee exemption; emails from various authorities concerned about our data collection; and a $30,000 bill for an error in the data collection process.[2]

Third, we estimate the effect of assigning cases to judges of different races or genders *without* controlling for other judge characteristics. Almost universally, prior research on diversifying the federal bench tries to "isolate" the effect on outcomes of a judge being a member of a minority group or being a woman. They do so by controlling for other judicial characteristics like Judicial Common Space scores, prior professional experience, law school attended, etc. But the judicial appointment process is highly political and strategic; nominees' identities play an important role in debates over their nominations, well before they are even seated on the bench. As a result, the women and judges of color that emerge from the appointment process may differ significantly from the White men that emerge from it. This means that many correlations between judges' races and/or genders and their other characteristics could be substantively important *consequences* of diversifying the bench. To draw lessons about whether diversity on the bench affects case outcomes, we do not want to treat those differences as inconveniences that need to be controlled for in a regression.

Most importantly, by not controlling for judicial ideology specifically, our empirical approach can detect effects that may be a consequence of the way that political actors strategically use di-

2. Thanks to the graciousness of one chief judge, we ultimately did not have to pay the $30,000 bill.

versity to achieve ideological goals in the judicial appointment process. It is well understood that ideological alignment is at the center of the appointments process—presidents and senators seek to appoint judges who share their ideological leanings. But racial and gender diversity has also been salient in the appointments process. How might these salient features in the appointment process—ideology, race, and gender—combine to affect who is ultimately appointed to the bench? By not controlling for judges' political ideologies, our analysis allows us to detect the ideological valence of the decisions made by appointees of different races or genders. In this respect, our analysis speaks to a recent literature showing how various types of selection effects explain documented gender and racial differences in both the composition and behavior of public officials (see, for example, Anzia and Berry 2011; Butler and Preece 2016; Fields 2016; Teele, Kalla, and Rosenbluth 2018; Broockman and Soltas 2020; Bernhard and Benedictis-Kessner 2021; Folke, Rickne, and Smith 2021).

By not controlling for other judicial characteristics, we also hope to alleviate some ambiguity that surrounds discussions of diversity in judging. Conditional effects (e.g., women make more liberal decisions in sex discrimination cases, after controlling for ideology and other factors) are often reported as overall effects (e.g., women make more liberal decisions in sex discrimination cases). Our aim is to estimate clearly communicable effects: the effect on case outcomes, within each party, of assigning cases to female judges or judges of color rather than White male judges.[3]

Our analysis is one of the largest empirical studies of federal cases to date. We collected and analyzed an original dataset of all civil rights cases filed in 20 federal district courts over multiple decades, totaling around 260,000 cases heard by 545 federal judges. The 20 district courts in our dataset are among the largest and most impactful in the nation. Combined, they have jurisdiction over 40% of the U.S. population, seat 40% of the federal district judges, and resolve 40% of federal civil rights lawsuits. We further supplement our analysis with a second dataset that includes every

3. Our approach to studying diversity on the bench has the additional benefit of avoiding the methodologically fraught exercise of trying to causally estimate the effect of judges' races or genders (see Greiner and Rubin 2011; Sen and Wasow 2016).

civil rights case filed in all federal district courts over two years.[4]

We focus on civil rights lawsuits because it is one of the most impactful areas of federal law. Many of the cases in our dataset were brought under some of the nation's landmark civil rights laws, such as the Civil Rights Act of 1964 and the Americans with Disabilities Act of 1990. At a more practical level, civil rights cases have been the focus of many prior studies on judicial diversity, and they also allow for more straightforward interpretations of effects in ideological terms: on average, outcomes favoring plaintiffs are "liberal" and outcomes favoring defendants are "conservative."[5]

We statistically test whether the judges of color and women appointed by presidents of each party, whom we collectively term "nontraditional appointees," cause different case outcomes relative to the White men appointed by presidents of the same party, whom we term "traditional appointees." There is substantial variation in the terms that prior scholars use to distinguish White and/or male judges from judges of color and women. We choose to borrow our terminology from Haire and Moyer (2015), one of the most widely cited studies on judicial diversity.

In contrast to the conventional wisdom, we do not find that nontraditional judges generate more liberal outcomes in an identity-salient area of law—civil rights cases. Among Democrats, we find no differences in average outcomes between cases assigned to nontraditional appointees and cases assigned to traditional appointees. Among Republicans, nontraditional appointees cause more conservative outcomes: they cause fewer settlements than traditional appointees, by approximately 2.2 percentage points, and more defendant wins, by approximately 1.2 percentage points. We confirm these findings with a supplemental analysis of a nationally representative dataset. For context, these within-party differences among Republican appointees are in the same direction and slightly larger than the effects we estimate for assigning cases to Republican rather than Democratic appointees. We do not find strong evidence that any of the effects differ among specific subsets of

4. The court records required to construct this dataset were provided by the Systematic Content Analysis of Legal EventS (SCALES) Open Knowledge Network (see https://scales-okn.org/).

5. As we discuss below, a hand-coding of a random sample of complaints shows that only a very small percentage of plaintiff claims could plausibly be characterized as advancing conservative causes (e.g., a White employee suing for race discrimination).

nontraditional appointees, such as Black judges, Latino judges, White women, etc.

We also test whether particular types of plaintiffs benefit from assignment to a nontraditional appointee. A commonly made claim is that judges of color will generate more favorable outcomes for people of color and female judges will generate more favorable outcomes for women (see Shayo and Zussman 2011; Harris and Sen 2019). We examine this issue by looking to see if judges of color and female judges cause different outcomes in cases filed by plaintiffs of color and female plaintiffs, respectively. We do not find evidence of substantial differences.

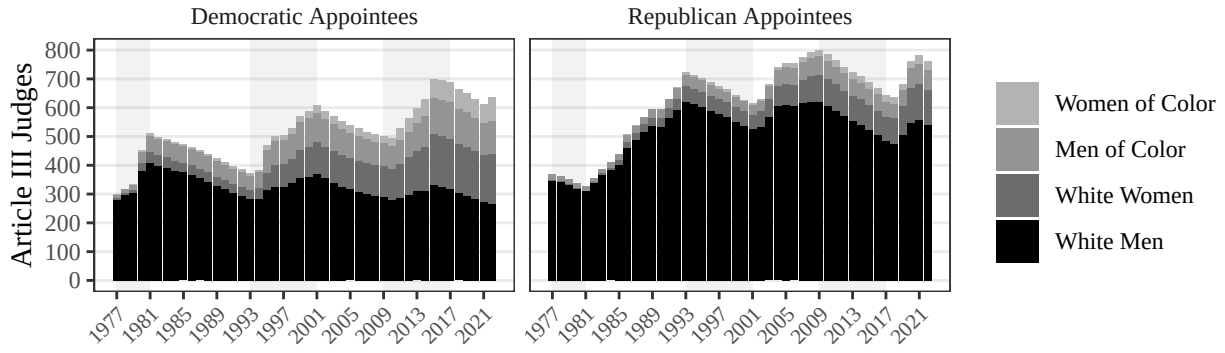
Admittedly, our results do not provide direct causal evidence of the historical effect of diversifying the federal bench. For example, the increasing presence of nontraditional appointees could influence the decision-making of their White male colleagues. Nor do we know the counterfactual decisions of the White men who would have been appointed had presidents decided to appoint fewer judges of color and women. However, the results do suggest that the historical appointment process created a situation in which Republican-appointed women and judges of color resolved cases more conservatively than Republican-appointed White men, whereas Democratic appointees resolved cases the same way regardless of their races or genders.

We theorize that the difference between the appointees of the two parties is explained by asymmetry in Republican and Democratic preferences for diversity. Since presidents must get their nominees confirmed by the Senate, we argue that the empirical pattern is consistent with a standard theory of bargaining over nominees in which Republican politicians place less importance on diversity on the bench than Democratic politicians do. According to the theory, this kind of preference asymmetry allows Republican presidents to "trade diversity" in exchange for nominating women and judges of color who will make more conservative decisions. We outline the logic of this theory below and analyze a formal model in Online Appendix D.

Racial and Gender Diversity on the Federal Bench

Historically, the federal judiciary has been composed of mostly White men. In fact, the first woman was not appointed to a federal Article III court until 1934 (Florence E. Allen), and the first Black judge was not named until 1950 (William Henry Hastie). Since the Carter administration, presidents have made concerted efforts to appoint more women and judges of color. We show these trends in Figure 1.

Figure 1: *Each panel plots a bar chart showing the total number of active and senior Article III federal judges serving on January 1 of each year, broken down by the party of the appointing president as well as the race and gender of the judges. Years in which Democratic presidents were appointing judges have a light gray background.*



What role do the racial or gender identities of judges play in how cases get resolved? Examining a wide range of civil cases heard in the federal courts of appeals, studies have coalesced around the notion that—after controlling for some other observable judicial characteristics—a judge's gender and race are associated with different outcomes in cases related to "racialized" or "gendered" issues, like employment discrimination (Morin 2014; Songer, Davis, and Haire 1994; Farhang and Wawro 2004), affirmative action (Kastellec 2013), and sex discrimination (Boyd, Epstein, and Martin 2010).[6] These studies are unified by their common conclusion that women or racial minorities induce more liberal (pro-plaintiff) outcomes, but there are a few exceptions (e.g., Morin

6. In an interesting twist, Glynn and Sen (2015) determine that, among appellate judges, having a daughter induces a man to decide cases in a more "feminist" direction in gender discrimination cases than those men without daughters.

2014 on Latino judges in employment discrimination cases, and Songer, Davis, and Haire 1994 on female judges in obscenity and search and seizure cases).

At the trial court level, an early contribution by Ashenfelter, Eisenberg, and Schwab (1995) did not find evidence that, controlling for a host of other factors, women resolve civil rights cases differently than men. Most recently, studies have concluded that cases assigned to women in the district courts are—adjusting for many other factors—more likely to result in pro-plaintiff outcomes in sex/pregnancy discrimination cases (Boyd 2016) and settled more frequently in civil rights cases (Boyd 2013). Controlling for other factors, Black judges increase pro-plaintiff outcomes in race discrimination suits (Boyd 2016), and judges of color are more likely than White judges to produce pro-claimant outcomes in Social Security disability cases (Boyd and Rutkowski 2020).

Why might the races or genders of judges affect case outcomes?[7] Scholars have focused on three main theories. The first posits that women and people of color have "different voices" they bring to the bench, leading them to resolve cases differently. The second theory holds that women and people of color—due to their different experiences in life—bring different information to judging. Finally, the third theory says that female judges and judges of color will act as substantive representatives of women and people of color, respectively, advocating for those groups' interests in their judicial decision-making. These theories are summarized in several recent papers, so we refer readers to those for more detailed descriptions (Boyd, Epstein, and Martin 2010; Boyd 2016; Harris and Sen 2019).

This literature has provided a large set of influential empirical findings. But it has some limitations. In addition to a relative lack of focus on the district courts, prior studies use research designs that are vulnerable to post-treatment statistical bias and that may partly obscure differences between judges of different races and genders. They also limit researchers' ability to detect selection effects.

7. Some have also theorized that judges' races and genders should not affect case outcomes since judges are all similarly socialized by the legal profession and should thus approach cases similarly (see Boyd, Epstein, and Martin 2010).

Common Research Designs Introduce Post-Treatment Bias

Federal court cases are typically assigned to judges randomly. Barring any flukes in random assignment and assuming that researchers account for the structure of the random assignment in the estimation process, this institutional feature provides an opportunity for a "natural experiment" that ensures differences in outcomes across judges are due to genuine differences between judges and not simply that different judges are assigned different kinds of cases. Unfortunately, many prior studies use research designs that do not exploit this random assignment and are vulnerable to statistical bias.

Most notably, many studies present statistical analyses that condition on post-treatment variables (for discussions of post-treatment bias, see Rosenbaum and Rubin 1985; Montgomery, Nyhan, and Torres 2018; Knox, Lowe, and Mummolo 2020). The most common way this manifests in courts research is when a researcher performs statistical analyses on subsets of cases that end in certain ways. For example, both Chew and Kelley (2009) and Collins, Manning, and Carp (2010) analyze samples of cases with published opinions. This is problematic because judges choose whether to publish opinions. Different judges may have different proclivities toward publication that could be correlated with other case characteristics. But there are other sources of post-treatment bias, such as controlling for case-level variables that occur after judges are assigned to cases (e.g., whether a case yielded a published opinion).

Many studies of diversity make these research design choices for theoretical or conceptual reasons. For example, some exclude settled cases based on the idea that judges can only "cause" outcomes where judges issue judgments or orders that end cases. We think this is an overly narrow view of the ways that judges can cause cases to end differently. Assignment of cases to different judges may cause outcomes that do not involve an ultimate judicial decision. For example, judges can affect case settlements through direct pathways, such as putting pressure on parties to settle. But their impact may also be through more indirect pathways if litigants make strategic decisions based on which judge is assigned to their cases. For example, if a judge who is known (or even

simply believed) to be favorable to civil rights plaintiffs is assigned to a civil rights case and the defendant decides to offer a settlement to avoid having their case overseen by a hostile judge, then this is a causal effect of that judge having been assigned to the case.

The possibility of statistical bias is a serious concern that undermines the interpretation of estimated effects. If, for example, some judges induce more settlements because they are known to be favorable to plaintiffs, then dropping cases that are settled from a dataset will disproportionately drop cases heard by these pro-plaintiff judges. There is mounting empirical evidence that the concern over post-treatment bias is justified. Recent research on the U.S. Courts of Appeals demonstrates that characteristics of published opinions appear to be correlated with the partisan make-up of the panels issuing those opinions (Carlson, Livermore, and Rockmore 2020). In the context of district courts, Hübert and Copus (2022) show that subsetting to cases that end in certain ways biases the effect of judge partisanship toward zero, potentially causing scholars to mistakenly conclude that political ideology matters less in district courts than other levels of the federal judiciary.

Even when studies rely on the random assignment of cases (and do not select on outcome variables), they do not always use estimation strategies that account for the randomization process, nor do they present evidence that their estimation strategies successfully exploit random assignment (e.g., balance tests). Random assignment of cases typically occurs within divisions of a district soon after cases are filed, not—as is often assumed—within districts. This may be a serious oversight. For example, it is well known that litigants purposefully choose to file cases in different divisions within a district in order to increase the odds of getting a more favorable judge (Botoman 2018).

Trying to "Isolate" the Effect of Judges' Identity Features Hides Important Selection Effects

Most quantitative studies of diversity in judging are framed around theories of judicial behavior that seek to explain *why* men and women, as well as White judges and judges of color, might generate different case outcomes. We briefly described these above. A common goal of these studies

is to address potential confounding caused by other judge characteristics (like judges' political ideologies). To do so, they typically include sets of judge-level control variables in their statistical analyses in order to "isolate" the specific effect of judges' races or genders on case outcomes. In our review of prior studies, nearly all of them include judge-level control variables.[8]

There are several problems with this. First, there is a standard selection on observables critique one could make about studies that control for specific judge-level characteristics. It is very difficult, if not impossible, to estimate an unbiased effect of a specific characteristic that is likely to be jointly determined with many other unobserved characteristics. We do not explore this critique in detail here, since there is a robust methodological literature on the topic (see Greiner and Rubin 2011; Sen and Wasow 2016).[9] Moreover, since different studies control for different judge-level variables—usually some combination of a judge's race, gender, age, ideology, tenure, prosecutorial experience, law school, religion, and prior judicial experience—it's unclear how to compare effects across studies.

There is another problem: reporting conditional effects can conceal insights about the overall impact of diversity on the bench. Consider a hypothetical dataset that generates the summary statistics presented in Table 1. A researcher is interested in seeing whether men or women are more favorable to plaintiffs, but she also has an intuition (based on theory) that ideologically liberal judges tend to rule for plaintiffs more often than ideologically conservative judges. She accordingly decides any regression looking at the effect of judge gender on case outcomes should include a control variable for judges' ideologies.

An analysis like that would find that women are 5 percentage points more likely to rule for

8. Some of these papers report descriptive (or "naive") estimates without judge-level controls, but it is unclear whether these estimates are intended to have causal interpretations. For example, reporting the difference in pro-plaintiff rates between men and women does not provide an "effect" of gender without some effort to exploit random assignment (e.g., include division-year fixed effects) or otherwise ensure that case characteristics are similar across treatment and control groups (see our discussion above).

9. There is also a compelling conceptual critique of this style of analysis. It implies that there is something like an essence of race or gender that can be empirically discovered once one controls away various other factors that, in reality, may be part of the complex social construction of a person's racial or gender identity.

Table 1: *Each cell contains the proportion of pro-plaintiff decisions for each combination of the two judge-level characteristics in a hypothetical dataset.*

	Men	Women
Ideologically Liberal	60%	65%
Ideologically Conservative	25%	30%

plaintiffs than men, holding constant judges' ideologies. With this finding in hand, suppose that a hypothetical paper's abstract reports: "Assignment of cases to women causes those cases to end in a pro-plaintiff way more often than assignment to men." However, this statement may be misleading. Suppose now that the women in this dataset are more likely to be ideologically conservative than men in this dataset. Depending on how different the pools of men and women are, it is possible that the assignment of cases to women actually decreases the rate of pro-plaintiff decisions![10]

This is an example of Simpson's Paradox, and it highlights the pitfalls of making claims about overall effects based on conditional estimates. This is substantively meaningful in our context: to know whether diversity on the federal bench affects case outcomes one would, at a minimum, want to know the overall effect. Yet the overall effect may be different in magnitude (or even direction) than a conditional effect. This is because differences in other judge characteristics might be an important *consequence* of diversifying the bench. Indeed, our own analysis demonstrates that nontraditional Republican appointees generate more conservative case outcomes than traditional Republican appointees, teaching us something important about the kinds of judges Republican presidents have put on the bench. Had we controlled for judges' ideologies, we may have inadvertently hidden this finding.

An important clarification is necessary: controlling for judge characteristics is not the same thing as looking at treatment effects among subsets of judges.[11] For example, since politicians

10. To be more concrete, suppose that 20 percent of the women in the sample are ideologically liberal whereas 40 percent of the men are. Then, the pro-plaintiff rate for women is $0.65 \times 0.20 + 0.30 \times 0.80 \approx 37\%$ and for men it is $0.60 \times 0.40 + 0.25 \times 0.60 \approx 39\%$. Clearly, assignment to a woman would *not* increase the likelihood of a pro-plaintiff outcome in that situation.

11. At a more technical level, the former involves adding judge characteristic variables to a regression or matching algorithm, and the latter involves *interactions* between the main treatment variable and these additional judge characteristics or performing separate analyses on subsets of a dataset.

from the two parties seem to approach the issue of diversity on the bench differently, there is a clear conceptual rationale for performing within-party analyses by comparing nontraditional appointees with the White men appointed by presidents of the same party. This is what we do in our analysis. We are examining conceptually-grounded heterogeneous treatment effects, not trying to "control for" a judge's political ideology, as is the standard rationale for including this variable in other judicial politics research.

Prior research studies often include judge-level control variables in order to test various hypotheses derived from theories of judicial behavior, and especially theories of race and gender in judging. Our focus is different since we seek to better understand whether the creation of a more diverse judiciary has affected case outcomes. As a result, we have no clear reason to include judge-level controls in our statistical models. But even if we were examining such theories of judicial behavior in this article, our core point still applies: including judge-level controls would weaken our ability to make accurate causal claims about the impact of diversity on the bench. When one includes judge-level controls in their main analyses, this is akin to skipping past the question "Do these judges have an effect?" and jumping straight to the question "*Why* do these judges have an effect?"[12] We think it is important to convincingly estimate unbiased effects before exploring the mechanisms driving those effects (see also p. 173 of Friedman et al. 2020).

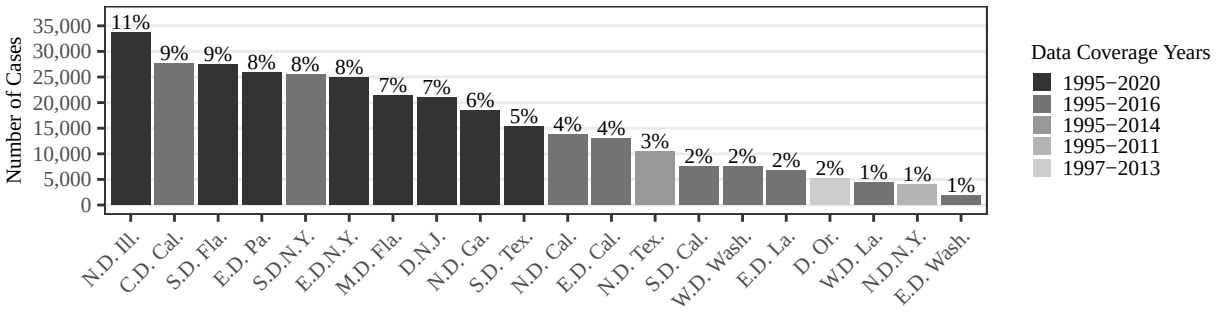
Data and Research Design

We constructed an original dataset of civil rights cases filed in 20 federal district courts. For most of these courts, our dataset includes all civil rights cases either filed between 1995 and 2016, or filed between 1995 and 2020. However, for three of the smaller courts, our dataset spans fewer years. Because our identification strategy requires us to estimate effects within district and within year,

12. We think this is, at least partially, due to an approach toward research that emphasizes deriving and articulating hypotheses from theory which can be "tested" with data. Our approach is *design based* (in the causal inference sense, see Angrist and Pischke 2010); we focus on estimating unbiased causal effects and then follow up with a novel theoretical framework that helps explain how those effects could have arisen (see below).

the slight differences in year coverage across the courts in our dataset do not create methodological problems. In Figure 2, we show the composition of our dataset.

Figure 2: *We show the number and percentage of cases in our dataset drawn from each of the 20 courts included in our analysis. For each court, we also use color shading to indicate the year range for which we have data from that court.*



Our dataset contains information about each case's characteristics as well as the presiding judge. Most of our case-level variables are drawn from the publicly available version of the Federal Judicial Center's Integrated Database (known as the "IDB" and available at https://www.fjc.gov/research/idb). We identified civil rights cases using the nature of suit (NOS) code variable in this dataset (i.e., those with NOS codes beginning with 44). In Online Appendix A, we provide additional information about these cases, both from the IDB and from a hand-coded random sample of cases in our dataset.

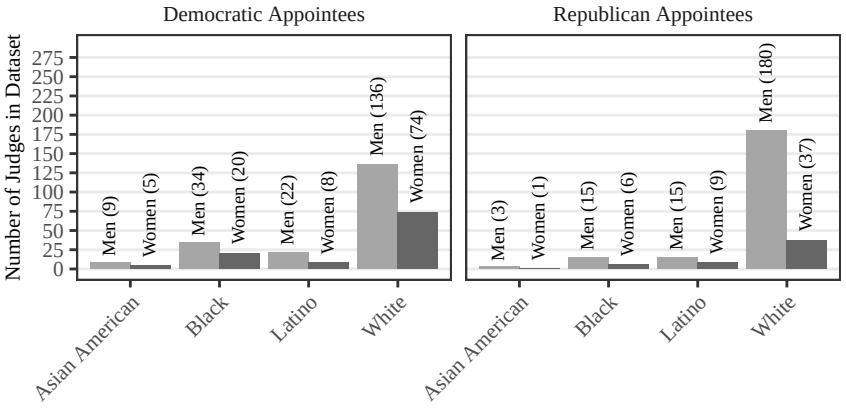
The publicly-available version of the IDB redacts the judge name from each case and only contains rudimentary information about each case's litigants. We add information about each case's litigants and presiding judge from an original database of docket sheets that we collected from the federal courts' fee-based online records system (called PACER) in connection with our ongoing research on judicial decision making. Twenty district courts from six different circuits and in four different regions of the U.S. issued us fee waivers that enabled us to access PACER for free. After more than a year of data collection and data cleaning, we were able to link judge-identifying information from these court records to the IDB.

Biographical data for each judge appearing in our dataset—and most importantly, each judge's race and gender—is drawn from the FJC's Biographical Directory of Article III Federal Judges (https://www.fjc.gov/history/judges). We follow prior research on federal judges and take the FJC's Biographical Directory as an authoritative source of information about each judge's racial and gender identity. We merge our three data sources together using both case numbers and judge names.

We supplement our main analysis with an analysis of a second dataset that contains all civil rights cases filed in every federal district court in 2016 and 2017. The docket sheets for this dataset were graciously provided by the Systematic Content Analysis of Litigation EventS (SCALES) Open Knowledge Network, which is working to build a platform that will provide open access to federal court data. We constructed this second dataset, which we call the SCALES dataset, using the same steps as we took for our main dataset.

Our statistical analyses focus on estimating the effect that assigning cases to nontraditional appointees instead of traditional appointees has on the outcomes of civil rights cases. Because our research design allows us to approximate a randomized experiment, we often describe the assignment of cases to nontraditional appointees as the "treatment" and the assignment of cases to traditional appointees as the "control." We also decompose this treatment variable and show separate estimates for several subgroups of nontraditional appointees. Our choice to define "treatment" and "control" in this way is simply a matter of labeling; our results would be identical (but with the opposite sign) if we reversed this. In Figure 3, we provide more detail about the racial and gender breakdown of the judges in our main dataset. In Online Appendix A, we provide this same information for the SCALES dataset.

Figure 3: *We plot the number of judges in our main dataset, broken down by judges' races, genders, and partisanship.*



District court cases can end in many different ways. In our analysis, we focus on the two most prevalent case outcomes in our dataset: settlements (45% of cases) and defendant wins (i.e., both involuntary dismissals and judgments favoring the defendant, together comprising 33% of cases). We code these case outcomes using both the IDB and our database of docket sheets. Online Appendix A provides additional descriptions of these outcomes and additional details about our coding process.[13]

Civil rights cases have a relatively clear political directionality to them (at least on average) since they almost always involve a plaintiff alleging a civil rights violation by a defendant.[14] We manually reviewed a random sample of cases and found that in only 3% of them could the plaintiff's legal claims plausibly be classified as ideologically conservative (e.g., a claim that an employee was discriminated against on the basis of being White). As a result, we interpret defendant wins as more "conservative." We interpret settlements as more "liberal" relative to defendant wins.[15]

13. To briefly summarize, we start with the IDB's "DISP" and "JUDGMENT" variables, and then use our docket sheet database to identify and recode case outcomes that are known to be systematically miscoded in the IDB (see Hadfield 2004).

14. For example, using the logic of the case space (see Lax 2011), a judge with a more liberal interpretation of discrimination law may be inclined to rule in favor of plaintiffs who present evidence that would be insufficient to convince a judge with a more conservative interpretation of discrimination law.

15. The latter interpretation is aided by the fact that we observe settlements and defendant wins trading off for one another in our main analysis.

In order to provide unbiased estimates of the effect of assigning nontraditional appointees to cases, we rely on the assumption that judges are randomly assigned to cases. However, for this to be a reasonable assumption, we take several steps that we outline in detail in Online Appendix B. In brief, we first drop subsets of cases (based on pre-treatment characteristics) that we suspect are not randomly assigned, and then we perform an aggressive statistical test of case randomization.

After this data cleaning, our randomization test provides convincing statistical evidence that the remaining cases in our dataset were randomly assigned to judges within each district's division and year. We can thus conceptualize our research design as a blocked (natural) experiment with as-if random treatment assignment within district-division-year randomization blocks. We use regression adjustment to account for the district-division-year randomization blocks, using the strategy described in Lin (2013) and research design recommendations provided in Lin, Green, and Coppock (2016).

Because our effects are only causally identified within each randomization block, we can only include randomization blocks that have sufficient variation in the treatment and control variables. This means our sample size varies across our analyses. We conduct our main analysis on 264,889 civil rights cases heard by 250 nontraditional appointees (164 appointed by Democratic presidents and 86 appointed by Republican presidents) and 295 traditional appointees (131 appointed by Democratic presidents and 164 appointed by Republican presidents). In all our figures, we include the sample size and number of judges for each analysis. We limit our analysis to cases assigned to judges appointed by Presidents Carter through Obama.

This estimation strategy allows us to recover credible causal estimates of the effect of assigning cases to nontraditional appointees, relative to assigning cases to White men. As we note above, we estimate effects separately for Democratic and Republican appointees to allow for the possibility that the effect of the assigned judge's race or gender on case outcomes matters differently depending on whether he or she is a Democratic or Republican appointee.

The regression model we use for each of our analyses is:

$$Y_i^p = \alpha^p + \beta^p \cdot N_i^p + \sum_{dy} \left\{ \phi_{dy}^p \cdot X_{idy}^p + \gamma_{dy}^p \cdot N_i^p \cdot (X_{idy}^p - \overline{X}_{dy}^p) \right\} + \varepsilon_i^p$$

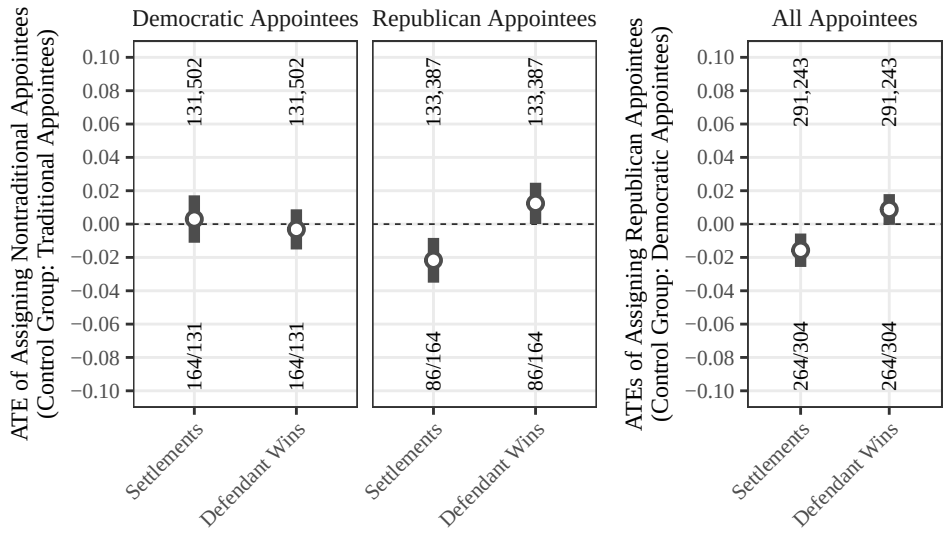
where i indexes cases, dy indexes a court division and case filing year (i.e., a randomization block), and p indexes the party of the appointing president. The variable N_i^p takes a value of 1 if case i is assigned to a nontraditional appointee and 0 if it is assigned to a traditional appointee. X_{idy}^p is a dummy variable indicating whether case i is in randomization block dy, and $\overline{X}_{dy}^p$ is the proportion of cases in our sample heard within randomization block dy.

Our main estimate of interest for each analysis is the estimate for β^p, which gives the average treatment effect of a case being assigned to a nontraditional appointee of party p. We cluster standard errors at the judge-level. We use the `estimatr` library for the R statistical programming language to estimate effects and standard errors (Blair et al. 2015).

Does Diversity on the Bench Benefit Civil Rights Plaintiffs?

We begin our analysis by estimating whether nontraditional appointees cause different case outcomes than traditional appointees. We plot these effects in the two left panels in Figure 4. Among Democratic appointees, there is no statistically significant difference in the average outcomes of the civil rights cases assigned to nontraditional appointees versus those assigned to traditional appointees. However, among Republican appointees, civil rights cases are less likely to end in settlements and more likely to end in wins for the defendant when assigned to a nontraditional appointee instead of a traditional appointee. Specifically, when a case is assigned to a nontraditional Republican appointee instead of a traditional Republican appointee, the case is 2.2 percentage points less likely to settle (p-value < 0.01) and 1.2 percentage points more likely to end in a judgment for the defendant (p-value < 0.05).

Figure 4: *Each point plots an average treatment effect on a specific case outcome (depicted on the x-axis), along with a 95% confidence interval using judge-clustered standard errors. In the left two panels, we plot our main effects. In the right panel, we plot the average treatment effect of assigning Republican appointees to cases (instead of Democratic appointees), which we provide for comparison. For each estimate, we present the number of cases in the analysis (top number) and the number of treatment/control judges (bottom number). Full results for this plot are available in Table E.1 in Online Appendix E.*



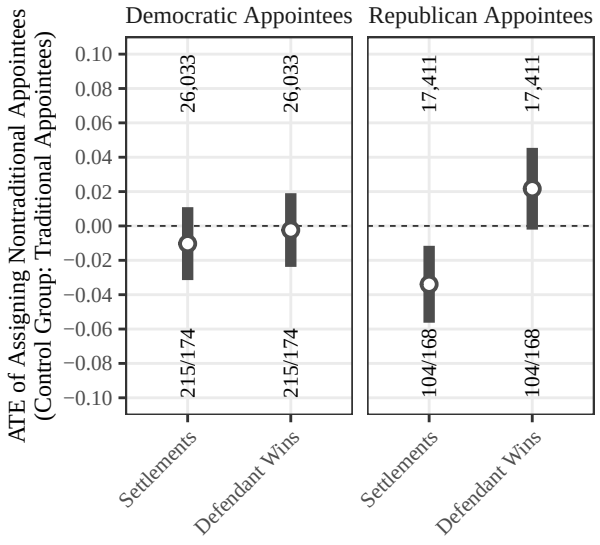
To place these effect sizes in substantive context, we also estimate partisan differences in case outcomes. Specifically, for each of the two outcomes, we estimate the average treatment effect of assigning cases to Republican appointees instead of Democratic appointees using the same estimation strategy as described above. In the right panel of Figure 4, we plot the estimates for these partisan effects. Assignment to a Republican appointee instead of a Democratic appointee decreases the probability of settlement by 1.6 percentage points and increases the probability of a defendant win by 0.9 percentage points (although the latter effect is not statistically significant at the 0.05 level). Our effects for nontraditional Republican appointees are both in the same direction and similar in magnitude to the effects for assignment to Republican appointees rather than Democratic appointees. This gives us further confidence that settlements can roughly be interpreted as more liberal outcomes in civil rights cases and defendant wins can roughly be interpreted as more

conservative outcomes in civil rights cases.

While these effect sizes may seem small in magnitude, it is well-known that many cases filed in federal district courts are frivolous. Effects are likely to be concentrated in the cases that are not frivolous, but identifying those cases with existing datasets and without introducing post-treatment bias is a challenge.[16]

Though our main dataset is expansive (covering 40% of the U.S. population), it is not a national sample. In order to mitigate concerns about generalizability, we supplement our main analysis with an analysis of the SCALES dataset, a smaller dataset but one that covers the population of civil rights cases filed in 2016 and 2017. Figure 5 displays the results, which support our previous findings: nontraditional Republican appointees cause fewer settlements and issue more decisions favoring defendants than traditional Republican appointees, while there is no difference among Democratic appointees.

Figure 5: *Each point plots an average treatment effect on a specific case outcome (depicted on the x-axis), along with a 95% confidence interval using judge-clustered standard errors. These analyses use the SCALES dataset. Full results for this plot are available in Table E.2 in Online Appendix E.*



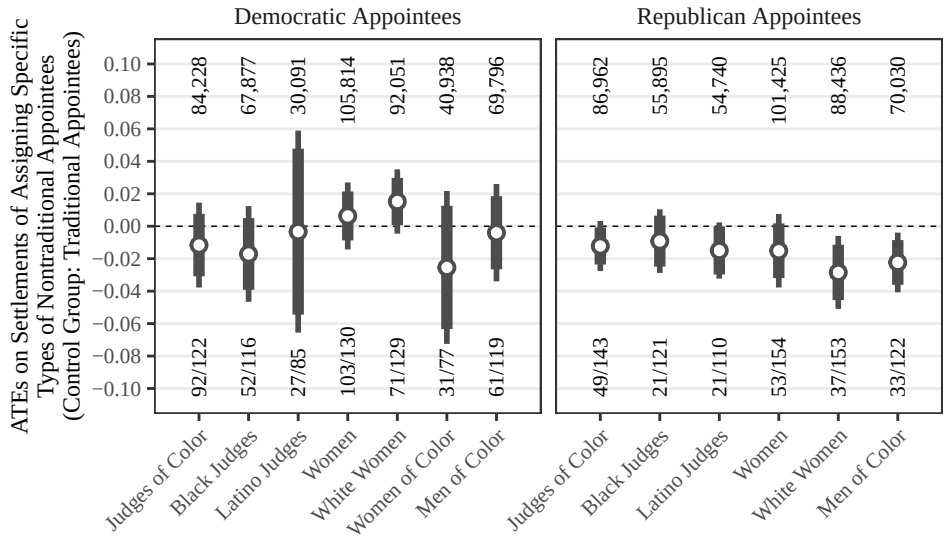
16. Many prior research studies attempt to focus on more "important" cases by, for example, analyzing only published opinions or cases that ended in a formal judgment. Unfortunately, this approach introduces post-treatment bias.

So far, we have only discussed whether cases end differently depending on whether they are assigned to nontraditional appointees or traditional appointees. However, we can also examine how the effects depend on which kind of nontraditional appointees are assigned to cases. In Figure 6, we show the effects on case outcomes for various subsets of nontraditional appointees. Each estimate is depicted with a circle and continues to use traditional appointees (i.e., White men) as the reference category. We restrict our analysis to only one outcome variable—settlements—but present analysis for defendant wins in Figure C.1 in Online Appendix C. Note that we only present estimates when we have at least 20 judges in the treatment group.

There is little evidence of effect heterogeneity by subgroup. Because of the large number of statistical tests we are conducting, we focus on the thinner, longer confidence intervals reported in the figure, which have been adjusted for multiple hypothesis testing.[17] We find a statistically significant effect only for Republican White women and men of color. Despite those significant effects, it is important to note that there are generally no substantial differences between subgroups of nontraditional appointees. As is apparent from the figure, the confidence intervals across the subgroups are largely overlapping, with most point estimates being covered. Moreover, if one wanted to explicitly compare these effects to one another (e.g., comparing the effect for women of color to the effect for White women), even the adjustments we make to the confidence intervals are not aggressive enough since they only account for each test independently. We would need to adjust for the many additional hypothesis tests that are implied by comparing the estimates with one another. We thus cannot conclude that the overall effects of nontraditional appointees are driven by a particular subset of those judges.

17. We adjust the confidence intervals using the Bonferroni method. Because the method will tend to overcorrect when applied to dependent hypotheses, we estimate the number of independent tests using the procedure described in Derringer (2018).

Figure 6: *Each point plots an average treatment effect on settlement, along with a 95% confidence interval using judge-clustered standard errors (the smaller bars present adjustments for multiple hypothesis testing using the Bonferroni method, with the number of independent tests estimated). Each estimate shows the estimated effect of assigning cases to judges with specific racial and/or gender characteristics, relative to traditional appointees. Full results for this plot are available in Table E.3 in Online Appendix E.*



Does Diversity on the Bench Benefit Women or People of Color?

The significance of diversity on the bench may extend beyond its general impact on case outcomes. In particular, nontraditional appointees might improve outcomes specifically for women and people of color. As Harris and Sen (2019) points out, "[r]esearch suggests that more women on the courts would lead to more decisions favorable to women [and] more people of color on the courts would lead to more decisions favorable to people of color..." (p. 243).

We use our dataset to explore this issue. First, we use standard automated methods to (1) identify which plaintiffs in our dataset are human individuals, and (2) predict the race/ethnicity and gender of those individuals. We use the `gender` package in R to predict each plaintiff's gender and the `wru` package in R to predict each plaintiff's race or ethnicity.[18] Given that data on the racial and

18. Though it is standard in the literature, the computational approach we use to predict the plaintiffs' gender and race/ethnicity is imperfect. To the extent that our automated process misclassifies some of the plaintiffs, this will introduce measurement error into our estimates. Although that error is not correlated with the assigned judge, these

gender identities of civil plaintiffs are not typically collected by federal courts, researchers must rely on these cutting-edge tools to predict these characteristics based on available data. These tools are now commonly used in political science (e.g., Grumbach and Sahn 2020; Grumbach, Sahn, and Staszak 2022). We discuss our data coding process in more detail in Online Appendix A. Using our predictions, we identify cases where we predict that all plaintiffs were people of color or White and cases where we predict that all plaintiffs were women or men. In the subsets of cases filed by plaintiffs of color and White plaintiffs (of any gender), we test if judges of color cause different outcomes than traditional appointees. In the subsets of cases filed by women and men (of any race), we test if female judges cause different outcomes than traditional appointees.

We only examine whether our effects vary by the identity of plaintiffs. In the context of Israeli small claims courts, Shayo and Zussman (2011) show that *defendant* identity also affects case outcomes. This is a less salient issue in our setting since the majority of the defendants in our dataset are organizations and governments. Indeed, only a small subset of the cases in our dataset (around 7%) feature individual (human) plaintiff(s) suing individual (human) defendant(s).

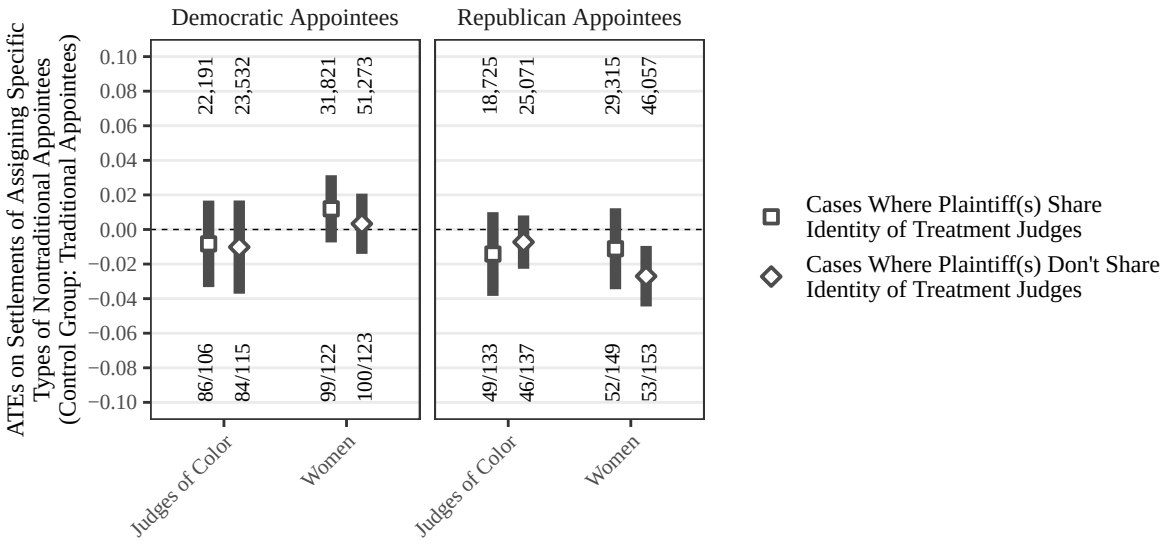
In this analysis, of the judges included in the "judges of color" category for Democratic appointees, 56% are Black, 30% are Latino, and 16% are Asian American.[19] Of the judges included in the "judges of color" category for Republican appointees, 43% are Black, 49% are Latino, and 8% are Asian American. Among the cases filed by plaintiffs of color who were heard by judges of color, 46% were filed by plaintiffs of the same race as the judge. Among the cases filed by women who were heard by female judges, 44% were filed by plaintiffs of the same race as the judge.

We plot the results of these analyses in Figure 7. First, we do not find evidence of any benefit to plaintiffs of color for having their cases assigned to a judge of color instead of a traditional

algorithms are known to have particular difficulty distinguishing White and Black names. The lack of significant effects on groups that include Black plaintiffs could thus be due to measurement error. As a robustness check, we also present results using a different package for classifying plaintiffs' races, `predictrace`. See Figure C.3 in Online Appendix C.

19. This adds up to greater than 100% because some of the judges are mixed race individuals and counted in multiple categories.

Figure 7: *Each point plots an average treatment effect on settlement, along with a 95% confidence interval using judge-clustered standard errors. The squares show the effect of assigning cases to judges with specific racial and/or gender characteristics in cases where the plaintiffs share the identity of the treatment group appointees. The diamonds show the effect of assigning cases to judges with specific racial and/or gender characteristics in cases where the plaintiffs do not share the identity of the treatment group appointees. Full results for this plot are available in Table E.4 in Online Appendix E.*



appointee, nor do we find any benefit to female plaintiffs for having their cases assigned to a woman instead of a traditional appointee (the square shaped point estimates in the figure). Second, we do not find any evidence that the treatment effect for judges of color varies by whether the plaintiffs are people of color or White; nor do we find any evidence that the treatment effect for women appointees varies by whether the plaintiffs are women or men (comparing the square shaped point estimates with the diamond shaped point estimates).

Because we are subsetting to a smaller number of cases filed by specific kinds of plaintiffs, testing for more nuanced effects than what we present in Figure 7 would slice our data very thinly. (For example, there were only around 150 cases filed by Asian plaintiffs that could have been assigned to an Asian American appointee.) Our focus is on testing the more statistically tractable claim summarized in Harris and Sen (2019) that judges of color will produce better outcomes for

plaintiffs of color and female judges will produce better outcomes for female plaintiffs. Nonetheless, for curious readers, we present additional results in Figure C.2 in Online Appendix C. We do not find evidence that nontraditional appointees provide more favorable outcomes for plaintiffs who share their identities.

Are Republicans Trading Diversity for Ideology?

What could explain why nontraditional Republican appointees resolve cases more conservatively, while there are apparently no differences among Democratic appointees? We argue that the asymmetry between the parties in our results, as well as the direction of those results, is broadly consistent with a strategic logic of partisan bargaining over judicial nominations in which diversity plays a role in presidents' strategic calculations (see also Asmussen 2011). This strategic logic starts with the premise that a president requires some buy-in from politicians of the opposing political party in order to successfully appoint his or her judges. This premise is reasonable in our context since a president's judicial nominees must be confirmed by the Senate, and, for the judges appointed in our dataset, both Senate rules and political norms made confirmation difficult without some support from members of both parties.[20]

In Online Appendix D, we analyze a simple formal model of political bargaining over judicial nominations in which two political parties have preferences over both the ideology and demographic characteristics of a judge who is nominated to fill a judicial vacancy. A key parameter in the model is $b_i \in \mathbb{R}$, which is player i's payoff from the president appointing a nontraditional appointee instead of a traditional appointee. So, holding all else equal: $b_i > 0$ implies player i prefers a nontraditional nominee, $b_i < 0$ implies player i prefers a traditional nominee, and $b_i = 0$ implies player i is indifferent about whether the nominee is nontraditional or not.

20. Only 39 of the 545 judges in our sample received their commission after Senate Democrats changed Senate rules so that confirmation of district judges could not be subject to filibuster. Moreover, of the judges in our sample appointed before this rule change, only 32 received their commission during periods of unified government and a filibuster-proof majority in the Senate.

Importantly, the player who makes a nomination in the model (i.e., the party that holds the presidency) pays a political cost if it does not win approval from the other party.[21] If a nominating president wishes to avoid paying this political cost, then she must provide some concession to the opposing party. This yields the following main result. (See Online Appendix D for the formal analysis and proofs.)

Proposition 1 (Trading Diversity)**.** Let $b_O \in \mathbb{R}$ be the opposition party's payoff from the appointment of a nontraditional appointee. In the unique equilibrium of the model of judicial nominations characterized in Online Appendix D, then relative to a president's traditional nominees, her nontraditional nominees will be

- more ideologically congruent with her if $b_O > 0$.
- no more or less ideologically congruent with her if $b_O = 0$.
- less ideologically congruent with her if $b_O < 0$.

Assuming that our model provides a reasonable approximation of the parties' incentives when bargaining over judicial nominees, then an empirical implication of Proposition 1 is that observed effects among one party's appointees reveal information about the other party's preference for (or against) the appointment of nontraditional nominees.

In light of this strategic logic, our main empirical findings provide support for the notion that Democrats place substantial weight on appointing a more inclusive federal bench, while Republicans do not. As a result, Democratic presidents cannot gain any ideological advantage by strategically choosing the identity of their nominees since Republican politicians are fairly indifferent about the identities of judicial nominees. On the other hand, Republican presidents can use Democrats' preference for diversity to extract ideological concessions, using diversity in appointments as a

21. These costs vary. For example, the opposing party could filibuster a nominee (a relatively higher cost) or release embarrassing information about a nominee (a relatively lower cost).

tool to appoint more conservative judges who would potentially engender more opposition from Democrats were they White men. We term this phenomenon "trading diversity" since Republican presidents can trade diversity for ideology, nominating nontraditional appointees who will act more conservatively on the bench.

Is there corroborating evidence to support the idea that Democrats value diversity while Republicans do not? Recall Figure 1, which demonstrates that Democratic presidents have appointed a substantially larger percentage of traditional appointees than Republicans. Between 1977 and 2020, 48% of Democratic presidents' appointees were women or judges of color, whereas only 26% of Republican appointees were women or judges of color.

On its face, this descriptive statistic lends support for the key implication of our trading diversity argument—that Democrats value diversity more than Republicans. However, what if these patterns in appointments are simply an artifact of differences in the pools of potential appointees available to Democratic presidents and Republican presidents? Perhaps Republican presidents value diversity just as much as Democratic presidents, but they face much higher search costs when attempting to recruit appointees from underrepresented groups.[22] However, using the logic of our theoretical model, this is unlikely. If Republican politicians placed a premium on diversity, then by Proposition 1, we would expect to see nontraditional Democratic appointees making more liberal decisions than the Democratic-appointed White men. We do not see this pattern in our data.

A similar argument allows us to rule out an obvious alternative explanation for our finding that nontraditional Republican appointees make more conservative decisions: namely, that Republican presidents engage in taste-based discrimination against nontraditional appointees. Our trading diversity argument relies on Republicans being indifferent about diversity. If instead they were *hostile* to diversity, then this would (also) lead them to appoint especially conservative nontraditional appointees.[23] However, this is not consistent with our other findings. If Republicans were

22. See Proposition D.2 in Online Appendix D.

23. This uses the well-known logic of taste-based discrimination first articulated by Becker (1957), which in this context predicts that Republican presidents would need to get some extra benefit (by way of more conservative nontra-

biased against nontraditional appointees, then from Proposition 1 we would also expect that bias to result in more conservative nontraditional Democratic appointees since Democratic presidents would need to nominate more conservative nontraditional appointees in order to overcome Republicans' racial or gender bias.

Our trading diversity argument is static and implies that appointing presidents make nominations that reflect the strategic environment at the point in time when they select judges for the bench. That said, our empirical results pool judges across time, comparing cases heard by all judges of color and women with cases heard by all White men. As is apparent from Figure 1, the bench has become more diverse as time passes. Our own dataset, for example, starts in 1995 with 36% of judges being nontraditional appointees and ends in 2020 with 55% being nontraditional appointees. There is a risk that the differences (or lack of differences) we see between traditional and nontraditional appointees in our analysis are driven simply by the fact that we are comparing nontraditional appointees disproportionately appointed more recently to traditional appointees appointed longer ago.[24] If this were true, it would undermine our theoretical argument. To guard against this, we re-estimate our effects from Figure 4 controlling for each presiding judge's appointing president. Our effects are nearly identical, suggesting that our main effects are not driven by a time trend (see Figure C.3 in Online Appendix C).

One final alternative explanation is that nontraditional Republican appointees may have been disproportionately appointed in times when there were fewer political constraints facing Republican presidents. If this had been the case, then Republican presidents would have been able to appoint more conservative judges (who just happened to be judges of color or women) because they faced less opposition from Democrats in the Senate. However, historical data on the partisan make-up of the Senate suggests that Republican presidents appoint nontraditional appointees when they face *more* of a political constraint. In our sample, nearly 51 percent of the White men appointed by

ditional appointees) in order to be willing to overcome their racial or gender bias and appoint nontraditional appointees.

24. In our dataset, the median appointment year of the traditional appointees is 1993, whereas the median appointment year of the nontraditional appointees is 1999.

Republican presidents were confirmed by a Republican majority in the Senate. In contrast, of the judges of color and women appointed by Republican presidents, 62 percent were confirmed when Republicans were in the *minority* in the Senate, consistent with the idea that political vulnerability encourages Republican presidents to trade diversity in order to see their judges confirmed.[25]

One might also take issue with the notion that presidents and senators are able to clearly discern ideological differences among those in their pool of potential nominees (see, for example, Hofer and Achury 2021). For the logic of the model to work, presidents and senators must be able to predict which judges will make more or less ideological decisions. However, the logic of the model does not require that they perfectly predict this, only that these predictions are accurate on average.[26] Moreover, we think our empirical analysis provides evidence that they do. Among Republican appointees, the differences in outcomes across nontraditional and traditional appointees are similar to the difference between Republican and Democratic appointees. If, as there is little doubt, presidents and senators can discern the across-party differences that affect case outcomes, then it is reasonable that they could also discern the within-party differences that yield effectively equivalent effects.

We argue that the logic of trading diversity provides a compelling explanation for our empirical results. However, our model black-boxes some potential mechanisms that are still consistent with our overall argument. For example, we do not know whether Republicans purposefully choose especially conservative women and people of color from a broad range of potential Republican appointees, or instead whether the pool of potential judges of color and women available to Republican presidents is especially conservative to begin with. Even if the pool of potential nontraditional Republican appointees is especially conservative, this is insufficient for explaining how they end up on the bench. To the extent that Democrats in the Senate have any leverage over Republican

25. For all district judges appointed by Republican presidents from 1977 to 2016, around 51 percent of the White men were appointed when Republicans held control of the Senate, versus 42 percent of the judges of color and women.

26. A simple extension to the model could incorporate uncertainty over the nominee's ideology, but this would make the model more complicated without altering the core finding that, in expectation, a president can "trade diversity" for ideology when members of the opposing party care about diversity on the bench.

nominations, it is the Democratic preference for a more diverse judiciary that enables Republican presidents to appoint women and judges of color who make more conservative decisions.

Conclusion

In this paper, we study how racial and gender diversity in the federal district courts impacts case outcomes. To do so, we analyze an original dataset of around 260,000 civil rights cases decided by 545 district judges over multiple decades in 20 district courts. These districts have jurisdiction over 40% of the U.S. population and seat 40% of the federal district judges. We find that among Democratic-appointed district judges, case outcomes are not affected by the identity of the judge assigned; among Republican-appointed district judges, case outcomes are more conservative when assigned to women or judges of color. We confirm these results with a supplemental analysis of the population of cases over a two year period. We further do not find statistical evidence of substantial differences across different subgroups of nontraditional appointees, nor do we find evidence that judges of color or women resolve cases differently when the plaintiff shares their race or gender.

Our approach differs from prior research on the role of judges' identities in that we shy away from attempting to "isolate" the effect of a judge's race or gender on their decision-making. Not only is this isolation strategy a difficult causal enterprise, but it also obscures the impact of diversity on the bench. Indeed, our results provide evidence against the common narrative that appointing a more diverse bench leads to more liberal decisions in certain kinds of salient cases, like discrimination cases. It is possible that, had we tried to control for judges' political ideologies, we would not have learned this.

Our analysis allows us to shine a light on the political process around judicial nominations. In a world where Democrats prioritize both the ideologies and identities of nominees, whereas Republicans prioritize only their ideologies, we would expect Democrats to be willing to endure an ideological cost for their pursuit of diversity and Republicans to exploit that fact. Our results are

broadly consistent with this logic since Republican presidents appear to appoint more conservative women and judges of color whereas Democrats achieve no ideological gains based on the racial or gender identities of the judges they appoint.

While our analysis provides a methodological and substantive step forward in understanding how diversity on the bench affects case outcomes, there is much left to do. First, and most obviously, future research should examine whether these effects hold for other kinds of cases and in other courts and time periods. A major challenge in looking at other case types is figuring out how to code the directionality of case outcomes in a way that is substantively meaningful and interpretable. The biggest challenge for studying additional courts and time periods is that access to court data (especially data on judges) is currently highly restricted. Second, future research should develop and empirically explore other ways that diversity on the bench matters. Here, we only examine whether cases are resolved differently depending on the identity of the judge assigned. While this clearly speaks to the larger questions about the impact of diversity on the bench, there are other intriguing questions to be answered. For example, does changing the composition of the judiciary influence which plaintiffs file suit, or what kinds of laws are passed in the first place? Does the presence of a larger number of nontraditional appointees on the bench influence how other judges (namely, White men) resolve cases?

References

Angrist, Joshua D., and Jörn-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics." *Journal of Economic Perspectives* 24 (2): 3–30.

Anzia, Sarah F., and Christopher R. Berry. 2011. "The Jackie (and Jill) Robinson Effect: Why Do Congresswomen Outperform Congressmen?" *American Journal of Political Science* 55 (13): 478–493.

Ashenfelter, Orley, Theodore Eisenberg, and Stewart J. Schwab. 1995. "Politics and the Judiciary: The Influence of Judicial Background on Case Outcomes." *The Journal of Legal Studies* 24 (2): 257–281.

Asmussen, Nicole. 2011. "Female and Minority Judicial Nominees: President's Delight and Senators' Dismay?" *Legislative Studies Quarterly* 36 (4): 591–619.

Becker, Gary S. 1957. *The Economics of Discrimination.* Chicago, IL: The University of Chicago Press.

Bernhard, Rachel, and Justin de Benedictis-Kessner. 2021. "Men and Women Candidates Are Similarly Persistent After Losing Elections." *Proceedings of the National Academy of Sciences* 118 (26): 1–5.

Blair, Graeme, Jasper Cooper, Alexander Coppock, Macartan Humphreys, Luke Sonnet, and Neal Fultz. 2015. *estimatr: Fast Estimators for Design-Based Inference.* https://declaredesign.org/r/estimatr/.

Botoman, Alex. 2018. "Divisional Judge-Shopping." *Columbia Human Rights Law Review* 49 (2): 297–344.

Boyd, Christina L. 2013. "She'll Settle It?" *Journal of Law and Courts* 1 (2): 193–219.

Boyd, Christina L. 2016. "Representation on the Courts? The Effects of Trial Judges' Sex and Race." *Political Research Quarterly* 69 (4): 788–799.

Boyd, Christina L., Lee Epstein, and Andrew D. Martin. 2010. "Untangling the Causal Effects of Sex on Judging." *American Journal of Political Science* 54 (2): 389–411.

Boyd, Christina L., and Adam G. Rutkowski. 2020. "Judicial Behavior in Disability Cases: Do Judge Sex and Race Matter?" *Politics, Groups, and Identities* 8 (4): 834–844.

Broockman, David E., and Evan J. Soltas. 2020. “A Natural Experiment on Discrimination in Elections.” *Journal of Public Economics* 188:1–14.

Butler, Daniel M., and Jessica Robinson Preece. 2016. “Recruitment and Perceptions of Gender Bias in Party Leader Support.” *Political Research Quarterly* 69 (4): 842–851.

Carlson, Keith, Michael A. Livermore, and Daniel N. Rockmore. 2020. “The Problem of Data Bias in the Pool of Published U.S. Appellate Court Opinions.” *Journal of Empirical Legal Studies* 17 (2): 224–261.

Carp, Robert A., and C.K. Rowland. 1996. *Politics and Judgment in Federal District Courts.* Lawrence, KS: University Press of Kansas.

Chew, Pat K., and Robert E. Kelley. 2009. “Myth of the Color-Blind Judge: An Empirical Analysis of Racial Harassment Cases.” *Washington University Law Review* 86 (5): 1117–1166.

Collins, Paul M., Jr, Kenneth L. Manning, and Robert A. Carp. 2010. “Gender, Critical Mass, and Judicial Decision Making.” *Law & Policy* 32 (2): 260–281.

Derringer, Jaime. 2018. “A Simple Correction for Non-Independent Tests.” *PsyArXiv [Preprint],* 1–18.

Farhang, Sean, and Gregory Wawro. 2004. “Institutional Dynamics on the U.S. Court of Appeals: Minority Representation under Panel Decision Making.” *Journal of Law, Economics, & Organization* 20 (2): 299–330.

Fields, Corey D. 2016. *Black Elephants in the Room: The Unexpected Politics of African American Republicans.* Oakland, CA: University of California Press.

Folke, Olle, Johanna Rickne, and Daniel M. Smith. 2021. “Gender and Dynastic Political Selection.” *Comparative Political Studies* 54 (2): 339–371.

Friedman, Barry, Margaret H. Lemos, Andrew D. Martin, Tom S. Clark, Allison Orr Larsen, and Anna Harvey. 2020. *Judicial Decision-Making: A Coursebook.* St. Paul, MN: West Academic Publishing.

Glynn, Adam N., and Maya Sen. 2015. "Identifying Judicial Empathy: Does Having Daughters Cause Judges to Rule for Women's Issues?" *American Journal of Political Science* 59 (1): 37–54.

Greiner, D. James, and Donald B. Rubin. 2011. "Causal Effects of Perceived Immutable Characteristics." *The Review of Economics and Statistics* 93 (3): 775–785.

Grumbach, Jacob M., and Alexander Sahn. 2020. "Race and Representation in Campaign Finance." *American Political Science Review* 114 (1): 206–221.

Grumbach, Jacob M., Alexander Sahn, and Sarah Staszak. 2022. "Gender, Race, and Intersectionality in Campaign Finance." *Political Behavior* 44:319–340.

Hadfield, Gillian K. 2004. "Where Have All the Trials Gone? Settlements, Nontrial Adjudications, and Statistical Artifacts in the Changing Disposition of Federal Civil Cases." *Journal of Empirical Legal Studies* 1 (3): 705–734.

Haire, Susan B., and Laura P. Moyer. 2015. *Diversity Matters: Judicial Policy Making in the U.S. Courts of Appeals.* Charlottesville, VA: University of Virginia Press.

Harris, Allison P., and Maya Sen. 2019. "Bias and Judging." *Annual Review of Political Science* 22 (1): 241–259.

Hofer, Scott, and Susan Achury. 2021. "The Consequences of Diversifying the US District Courts: Race, Gender, and Ideological Alignment through Judicial Appointments." *Justice System Journal* 42 (3–4): 306–324.

Hübert, Ryan, and Ryan Copus. 2022. "Political Appointments and Outcomes in Federal District Courts." *Journal of Politics* 84 (2): 908–922.

Kastellec, Jonathan P. 2013. "Racial Diversity and Judicial Influence on Appellate Courts." *American Journal of Political Science* 57 (1): 167–183.

Knox, Dean, Will Lowe, and Jonathan Mummolo. 2020. "Administrative Records Mask Racially Biased Policing." *American Political Science Review* 114 (3): 619–637.

Labaton, Stephen. 1993. "Clinton May Use Diversity Pledge to Remake Courts." *New York Times,* accessed 2022-08-26. https://www.nytimes.com/1993/03/08/us/clinton-may-use-diversity-pledge-to-remake-courts.html.

Lax, Jeffrey R. 2011. "The New Judicial Politics of Legal Doctrine." *Annual Review of Political Science* 14:131–157.

Lin, Winston. 2013. "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique." *The Annals of Applied Statistics* 7:295–318.

Lin, Winston, Donald P. Green, and Alexander Coppock. 2016. *Standard Operating Procedures for Don Green's Lab at Columbia.* https://alexandercoppock.com/Green-Lab-SOP/Green_Lab_SOP.pdf.

Montgomery, Jacob M., Brendan Nyhan, and Michelle Torres. 2018. "How Conditioning on Post-treatment Variables Can Ruin Your Experiment and What to Do about It." *American Journal of Political Science* 62 (3): 760–775.

Morin, Jason L. 2014. "The Voting Behavior of Minority Judges in the U.S. Courts of Appeals: Does the Race of the Claimant Matter?" *American Politics Research* 42 (1): 34–64.

Rosenbaum, Paul R., and Donald B. Rubin. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score." *The American Statistician* 39 (1): 33–38.

Sen, Maya, and Omar Wasow. 2016. "Race as a Bundle of Sticks: Designs that Estimate Effects of Seemingly Immutable Characteristics." *Annual Review of Political Science* 19 (1): 499–522.

Shayo, Moses, and Asaf Zussman. 2011. "Judicial Ingroup Bias in the Shadow of Terrorism." *The Quarterly Journal of Economics* 126 (3): 1447–1484.

Songer, Donald R., Sue Davis, and Susan Haire. 1994. "A Reappraisal of Diversification in the Federal Courts: Gender Effects in the Courts of Appeals." *Journal of Politics* 56 (2): 425–439.

Teele, Dawn Langan, Joshua Kalla, and Frances Rosenbluth. 2018. "The Ties That Double Bind: Social Roles and Women's Underrepresentation in Politics." *American Political Science Review* 112 (3): 525–541.