Online Appendix for

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

Ryan Copus, UMKC School of Law
Ryan Hübert, UC Davis
Paige Pellaton, UC Davis

March 24, 2024

For Online Publication

Replication code and data will be made available.

Contents

A Our Dataset

We constructed a dataset of every civil rights case filed in 20 district courts over the course of multiple decades. We created the dataset using three sources: (1) the FJC's Integrated Database (https://www.fjc.gov/research/idb/), (2) an original database of docket sheets collected from PACER, and (3) the FJC's Biographical Directory of Federal Judges (https://www.fjc.gov/history/judges). We merge the first two data sources together using each case's docket number. We merged the last dataset using judges' names. From these data sources, we coded our main variables of interest.

Treatment variable The treatment variable in our main analysis is a binary variable indicating whether a judge is a "nontraditional appointee" or a "traditional appointee." In our dataset, traditional appointees are those whom the FJC's Biographical Directory of Federal Judges classifies as "White" in the *Race or Ethnicity* field and "Male" in the *Gender* field. Nontraditional appointees are all other judges.

We identified the presiding judge for each case from its docket sheet. At the beginning of each docket sheet, there is an "Assigned to" field. We accordingly refer to this as the "assigned judge." However, closer inspection of the docket sheets revealed that this field is updated whenever a case is reassigned to another judge. As a result, the assigned judge is the *last* judge assigned, not the first judge assigned to a case. Because we don't know why or how some cases are reassigned to different judges, we cannot be confident that the assigned judge in each case is randomly assigned. To get around this issue, we used automated methods to scan the docket sheet entries to identify the first judge to take any action in a case. We are sufficiently confident that this judge, who we call the "first judge," is the judge to whom the case is randomly assigned when it is filed. We manually coded a random sample of 200 cases and found that our automated method accurately identified the first assigned judge in 95 percent of the cases.

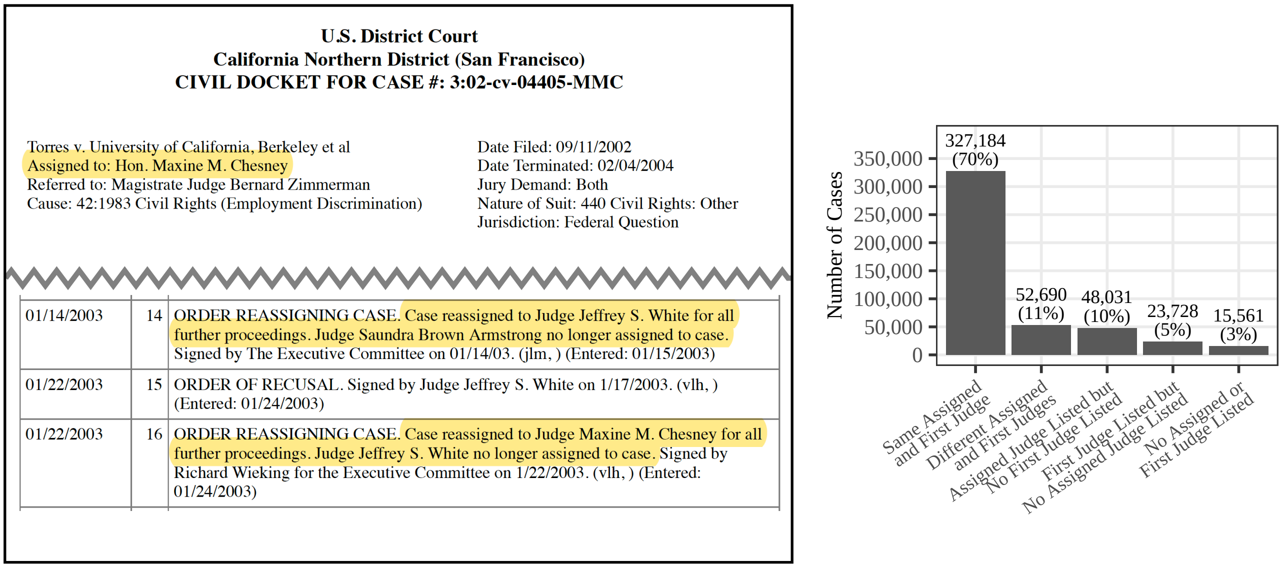
In the left panel of Figure A.1, we include a screen grab of a portion of the docket sheet corresponding to one of the cases in our dataset. The docket entries (below the jagged line) demonstrate that District Judge Saundra Brown Armstrong was initially assigned to the case. District Judge Maxine Chesney was eventually reassigned to this case, after which the "Assigned to" field was updated to reflect the reassignment. In our dataset, Judge Armstrong is coded as the first judge, and Judge Chesney is coded as the assigned judge. So, Judge Armstrong is the judge we use for our analysis.

We use the first judge to code our treatment variable. This means that for some set of cases, we coded a different judge than the one listed in the "Assigned to" field. Moreover, for another set of cases, the "Assigned to" field is blank even though our automated methods revealed that there is a first judge who was initially assigned to the case.[1]

In the right panel of Figure A.1, we depict the number of cases where the first and assigned judges

1. Our best guess for why this is the case is that when a judge leaves the bench or otherwise reduces their caseload, some cases may be taken off their docket but never reassigned because they are not currently live cases.

Figure A.1: *In the left panel, we provide portions of a screen grab of a docket sheet from a case in our dataset. It shows the kind of information available in federal district court docket sheets, including the identity of the presiding judge. Note that this case was reassigned from one judge to another. In the right panel, we plot the number of cases falling into one of four categories depending on the assigned and first judges.*



U.S. District Court
California Northern District (San Francisco)
CIVIL DOCKET FOR CASE #: 3:02-cv-04405-MMC

Torres v. University of California, Berkeley et al
Assigned to: Hon. Maxine M. Chesney
Referred to: Magistrate Judge Bernard Zimmerman
Cause: 42:1983 Civil Rights (Employment Discrimination)

Date Filed: 09/11/2002
Date Terminated: 02/04/2004
Jury Demand: Both
Nature of Suit: 440 Civil Rights: Other
Jurisdiction: Federal Question

01/14/2003	14	ORDER REASSIGNING CASE. Case reassigned to Judge Jeffrey S. White for all further proceedings. Judge Saundra Brown Armstrong no longer assigned to case. Signed by The Executive Committee on 01/14/03. (jlm,) (Entered: 01/15/2003)
01/22/2003	15	ORDER OF RECUSAL. Signed by Judge Jeffrey S. White on 1/17/2003. (vlh,) (Entered: 01/24/2003)
01/22/2003	16	ORDER REASSIGNING CASE. Case reassigned to Judge Maxine M. Chesney for all further proceedings. Judge Jeffrey S. White no longer assigned to case. Signed by Richard Wieking for the Executive Committee on 1/22/2003. (vlh,) (Entered: 01/24/2003)

are the same, where we used the first judge instead of the assigned judge (because they were different, or the assigned judge was missing), and the number of cases with neither an assigned nor first judge listed.[2]

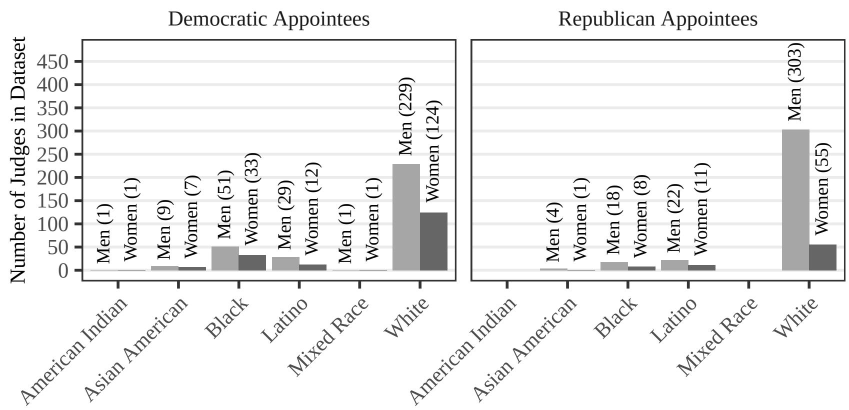
In the main text, we present descriptive statistics on the gender and racial breakdown of the judges in our main dataset. In Figure A.2, we present descriptive statistics on the judges in the SCALES dataset, broken down by the party of the appointing president as well as the judges' races and genders.

Outcome variable Case outcomes are coded using information from the FJC IDB, as well as from the cases' docket sheets. We primarily rely on the IDB for case outcomes, but prior research suggests that some outcomes in the IDB—and specifically voluntary dismissals—are miscoded (see Hadfield 2004). A large manual review of the IDB confirmed that many voluntary dismissals were systematically miscoded, as were many judgments for which the IDB did not identify a winning party. We briefly describe the problem and our solution.

The "voluntary dismissal" category in the IDB consolidates three substantively different types of outcomes: unilateral plaintiff withdrawals, joint withdrawals filed by both parties, and settlements. A plaintiff can only unilaterally withdraw their case (by way of a voluntary dismissal) before the

2. These are often cases that are only assigned to a magistrate judge or cases assigned to a "duty judge" who hears a large number of smaller cases. Different courts have different rules about these kinds of cases.

Figure A.2: *We plot the number of judges in the SCALES Dataset, broken down by judges' races, genders, and partisanship.*



defendant files an answer or motion for summary judgment (see rule Federal Rule of Civil Procedure 41(a)(1)(A)(i)). This means that if a plaintiff wishes to withdraw their case after an answer or motion for summary judgment is filed, they must get the defendant(s) to agree. This is especially important because many settlements *also* involve a pro forma notice of joint voluntary dismissal. So, joint voluntary dismissals can indicate that either the plaintiff is withdrawing their case or the case has been settled.

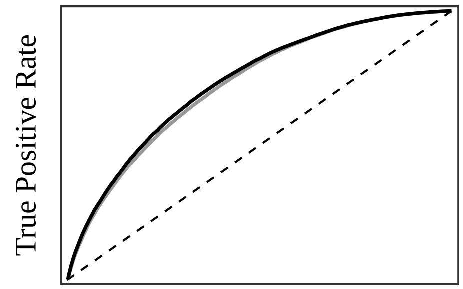
We use information available in the docket sheets to recode voluntary dismissals to capture this additional nuance. Specifically, for every case that the IDB classifies as a voluntary dismissal:

- If the docket sheet explicitly mentions a settlement occurred, we reclassify the case outcome to "settlement."
- If the docket sheet does not explicitly mention a settlement occurred then we reclassify the case outcome to "joint voluntary dismissal" if it either (1) mentions a joint voluntary dismissal or (2) was filed after the defendant filed an answer or a motion for summary judgment,
- If the docket sheet does not explicitly mention a settlement occurred, nor does it reference a joint voluntary dismissal, then we keep it as a "voluntary dismissal" which we assume is unilateral.

Finally, for any case coded as a judgment for an unknown party that explicitly references a settlement, we recoded that case outcome to "settlement."

As one final check, we looked to see whether the IDB's miscoding was correlated with either the party of judges' appointing presidents or whether they are nontraditional appointees. We do so by using aggressive machine learning algorithms to see if these judge characteristics are (partially) predictive of whether case outcomes are miscoded in the IDB. We use the same process as our randomization test described in Appendix B. We plot the ROC curves in Figure A.3, showing that judge characteristics do *not* help predict whether a case outcome is miscoded. This is evidence of classical (random) measurement error in our dependent variable, which will not systematically bias our estimates.

Figure A.3: *We used two ensemble machine learning algorithms to predict when case outcomes were miscoded. In the benchmark model, we only use case-level characteristics. In the saturated model, we include case-level characteristics plus assigned judge characteristics (i.e., whether judges were Republican appointees or nontraditional appointees). We plot ROC curves for both models, demonstrating judge characteristics provide no additional predictive power over case characteristics, strong evidence that the IDB's miscoded outcomes are uncorrelated with the assigned judge.*



We analyze average treatment effects for the top two outcomes in our dataset:

- *Settlements* (45% of our dataset):
  - All cases in which the IDB's DISP variable takes a value of 13 (case settled).
  - All cases in which the IDB's DISP variable takes a value of 2 (dismissal for want of prosecution), 3 (dismissal for lack of jurisdiction), 12 (voluntarily dismissed), or 14 (other dismissal), and the docket sheet explicitly mentions a settlement.
  - All cases in which the DISP variable takes a value of 4 (default judgment), 5 (consent judgment), 6 (judgment on motion before trial), 7 (judgment after jury verdict), 8 (judgment after directed verdict), 9 (judgment after court trial), 15 (judgment after award of arbitrator), 16 (stayed pending bankruptcy), 17 (other judgment), 18 (statistical closing), 19 (judgment after appeal of magistrate judge affirmed) or 20 (judgment after appeal of magistrate judge denied); the JUDGMENT variable takes a value of 0 (missing), 4 (unknown) or -8 (missing); and the docket sheet explicitly mentions a settlement.
- *Defendant wins* (i.e., involuntary dismissals or judgments for defendant, 33% of our dataset):
  - All cases in which the DISP variable takes a value of 2 (dismissal for want of prosecution), 3 (dismissal for lack of jurisdiction), or 14 (other dismissal).
  - All cases in which the DISP variable takes a value of 4 (default judgment), 5 (consent judgment), 6 (judgment on motion before trial), 7 (judgment after jury verdict), 8 (judgment after directed verdict), 9 (judgment after court trial), 15 (judgment after award of arbitrator), 16 (stayed pending bankruptcy), 17 (other judgment), 18 (statistical closing), 19 (judgment after appeal of magistrate judge affirmed) or 20 (judgment after appeal of magistrate judge denied), and the JUDGMENT variable takes a value of 2 (defendant win).

The remaining outcomes are: joint voluntary dismissals (7%), (unilateral) voluntary dismissals (4%), remands to state court (3%), judgments for an unknown party (3%), judgments for the plaintiff (3%), inter-district transfer (2%), remand to agency (0.1%).

Plaintiffs' races and genders In one of our analyses, we investigate whether nontraditional appointees cause different case outcomes based on shared racial or gender identities with the plaintiffs. The FJC does not report plaintiffs' genders or races, so we use automated techniques to predict plaintiffs' race and gender based on their name (extracted from each case's docket sheet) and their county of residence (extracted from the IDB). As some plaintiffs in our sample are government entities, businesses, or other organizations, we employ automated methods to exclude non-human plaintiffs. These methods have undergone thorough validation and utilize custom dictionary approaches that we vetted extensively.

We predicted the gender of each plaintiff using two methods: First, we utilized the `gender` package by Blevins and Mullen (2015) to infer gender based on historical data from the U.S. Social Security Administration. This method used a plaintiff's first name to classify them as "male" or "female" based on the likelihood a name was associated with a particular gender at a given point in time. For plaintiffs whose first names yielded no clear prediction, we reran the procedure using a plaintiff's middle name, if available. As a second strategy, we classified gender using the Integrated Public Use Microdata Series method in the `gender` package. The two classifications were in agreement for 91% of plaintiffs. In our analysis, we use the SSA method and supplement it with the IPUMS method when the SSA method yields no prediction.

To determine the race/ethnicity of plaintiffs in our sample, we used the `wru` package by Imai and Khanna (2016), which predicts a person's race based on their surname and geolocation (county). This package, used by others like Grumbach and Sahn (2020), uses the U.S. Census Bureau's Surname List (2000 version) and geocoded voter registration records to predict the probability that a plaintiff is White, Black, Hispanic, Asian, or Other. When a most-likely race could not be established, we used the `wru` package again to predict race based on surname only (i.e., without county). We used the resulting probabilities to code a prediction of the most-likely race of each plaintiff. Specifically, we compared the predicted probabilities for each race and coded a plaintiff's race when the probability they were one race (e.g., White) exceeded the combined probabilities of all other races (e.g., non-White). This was a conservative coding decision that ensures our analysis does not include cases with plaintiffs that `wru` cannot easily classify. As a second strategy, we employed the `predictrace` package developed by Kaplan (2021), which uses first names and surnames to predict the most prevalent race associated with each.

The `wru` and `predictrace` packages predict the same race for 58% of the plaintiffs in our dataset and a different race for 12% of the plaintiffs in our dataset. For the remaining plaintiffs: 27% were missing a prediction from `wru`, 3% were missing a prediction from `predictrace`, and 3% were missing predictions from both. Of the 11% of plaintiffs that were coded differently across the two packages, the vast majority of them were coded as White by `predictrace` and either Black or Hispanic by `wru`.

For the `wru` package (whose predictions we use in the main text), we summarize the distribution of predicted probabilities of plaintiff race in Table A.1. Specifically, for each plaintiff's racial classification (left column), we present the mean of the predicted probabilities (the five columns on the right). As expected from the prior literature, the algorithm has the hardest time distinguishing Black and White names.

Table A.1: *We summarize the distribution of predicted probabilities for the plaintiffs' race classifications. Note: the "None" category includes plaintiffs for whom the `wru` algorithm returned predicted probabilities less than 0.5 for every racial/ethnic category.*

Classified Plaintiff Race	Mean Pr(White)	Mean Pr(Black)	Mean Pr(Hispanic)	Mean Pr(Asian)	Mean Pr(Other)
White (40%)	0.750	0.086	0.039	0.013	0.113
Black (14%)	0.200	0.666	0.026	0.008	0.100
Hispanic (12%)	0.037	0.008	0.902	0.022	0.031
Asian (2%)	0.032	0.016	0.019	0.823	0.109
Other (1%)	0.198	0.113	0.070	0.031	0.589
No Dominant Prediction (23%)	0.348	0.288	0.105	0.031	0.228
None (8%)	—	—	—	—	—

Because litigant race/ethnicity is not provided to us via case filings, we cannot directly assess the accuracy of the `wru` and `predictrace` predictions in our sample. However, we conducted a robustness exercise by comparing our automated prediction method against the races/ethnicities of federal district judges appointed since President Carter, whose races are reported by the FJC. The classifications generated by `wru` matched the reported race for 78% of judges in the FJC database. Of the inaccurate predictions, the vast majority were Black judges that `wru` predicted were most likely White or White judges that were predicted as most likely Black, or White judges that we did not assign a most-likely race. The judge race predicted by `predictrace` matched the reported race for 86% of judges in the FJC database. The majority of inaccurate predictions were also Black judges that `predictrace` coded as White.

Using automated methods, like Imai and Khanna 2016's Bayesian Improved Surname Geocoding (BISG), to predict race/ethnicity based on names is a relatively new endeavor in the literature that has been enabled by the recent development of powerful statistical algorithms and large datasets. Despite their novelty, these methods, notably `wru`, have gained widespread acceptance in the literature. For instance, `wru` has been used by Abott and Magazinnik (2020) to identify Latino school board candidates and by Grumbach, Sahn, and Staszak 2022 to predict the race of campaign contributors. For a comprehensive overview of the increasing use of automated methods to predict race in political science and other disciplines, we refer readers to Clark et al. (2021). As for the overall accuracy of these classification methods, a recent research note by Rosenman et al. (2023) found that the `wru` classifier had just a 13.2% error rate when applied to a validated sample of 38 million voters. Thus, we rely on `wru` in the main text, and, as a robustness check on our main findings, `predictrace` in our online appendix (see Figure C.2).

Types of civil rights cases To gain insight into the types of civil rights involved in our case sample, we randomly sampled 100 original complaints in our data, read the full complaints, and categorized them by the type of civil rights discussed in each complaint. The distribution of our sample can be found in Table A.2.

Table A.2: *The percentage of suits belonging to each civil rights category based on a random sample of 100 original complaints. Percentages add up to greater than 100% because some of the cases involve more than one type of alleged civil rights violation.*

ADA Access	Employment	Race	Gender	Age	Disability	Police
15%	40%	18%	16%	11%	23%	20%

B Leveraging the Random Assignment of Cases to Judges

We rely on the random assignment of cases to judges in federal district courts to estimate causal effects. However, ensuring that we properly leverage random assignment requires some additional work, which we outline here.

Collect qualitative information about case assignment procedures We collected qualitative information from each court's General Orders and Local Rules relating to the processes used to assign cases to the judges in our sample. Based on our review of this information, we concluded that it's standard practice for cases to be randomly assigned to judges as they are filed and that this random assignment typically occurs within the division of the district in which the case was filed.[3]

Practically speaking, within a division and unit of time, we can consider all cases filed to be randomly allocated to the available district judges.[4] (Again, it is insufficient to assume random case assignment within each district.) We use the filing year as our unit of time. We refer to each district-division-year combination as a "randomization block."[5] In the main text, we discuss how we recover the random assignment by way of our statistical estimation model.

Address problems with judge assignment in some cases Each case in our dataset must have exactly one presiding district judge. We first drop all cases with no presiding district judge. Then

3. For example, the Northern District of California has four divisions (sometimes called offices, duty stations, or courthouses), based in Eureka, Oakland, San Francisco, and San José. A plaintiff can file their case in any of these divisions (see page 3 of https://cand.uscourts.gov/filelibrary/1243/Atty_Case_Opening_Guide_2019.pdf).

4. Semi-retired "senior judges" have more latitude over how many cases they are assigned but not *which* cases they are assigned. We find no evidence that cases are assigned non-randomly to these judges, so we include them in our analysis. For substantive reasons, we drop cases heard by judges from another court who are temporarily assigned to hear cases in a particular court.

5. Ideally we would use a finer unit of time, such as quarter, month, or week. However, the finer our measure, the more difficult it is for us to estimate effects since it will cause us to dramatically reduce our within-block sample size. Moreover, as we demonstrate below, cases appear to be as-if randomly assigned within each district-division-year.

we took the following additional steps:

Step 1: We dropped any case that we have reason to believe was not randomly assigned to a judge via the normal randomized procedure. This included: (1) cases that are classified as multi-district litigation cases; (2) cases that appear to be one of several "related cases," which are assigned to a specific judge because they were filed on the same day, in the same district-division, with the same nature of suit code; or (3) cases that were either post-appeal actions or appeals of a magistrate judge's decision.[6]

Step 2: We dropped any case that was heard by a judge who does not sit in the filing court (a "visiting judge") or any judge appointed by a president before Jimmy Carter or after Barack Obama. The justification for the first of these is that we do not want our effects to be influenced by judges who do not regularly sit in the district and who may only hear a limited number of cases. The justification for the second of these is that these judges do not hear a large number of cases. Neither of these decisions creates a problem for our causal identification strategy since the appointing president is a pre-treatment variable.

Step 3: We drop all cases heard within a specific court-division-year block in which fewer than five cases are heard by nontraditional appointees or fewer than five cases are heard by White men. We do this following the recommendations provided in Lin, Green, and Coppock (2016). This is a consequential data-cleaning step, as it forces us to drop a sizeable number of cases that are (in principle) randomly assigned to a judge. We also estimate our main effects with different minimum thresholds, one and ten, and our results don't change.

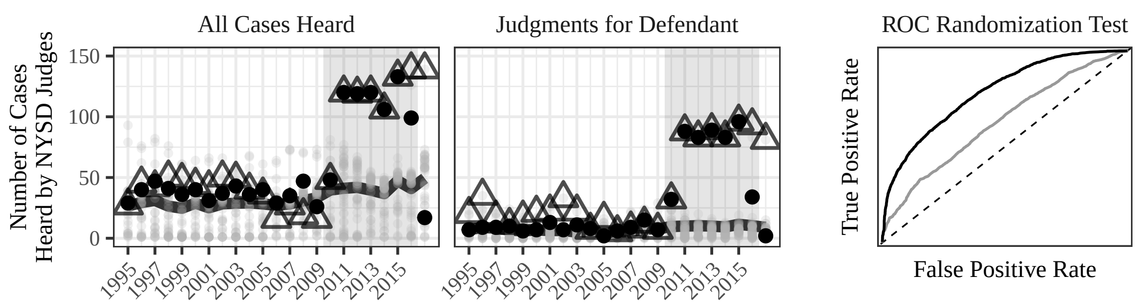
Provide quantitative evidence in support of our approach In our analysis, we assume cases are assigned as-if random conditional on court-division and case filing year. We conduct an aggressive test of this assumption using a computational approach described in Hübert and Copus (2022). First, we use a machine learning algorithm to predict unit-level treatment status using district-division-year randomization blocks, since our identifying assumption is that potential outcomes are independent of treatment only after conditioning on these randomization blocks. We call this our "benchmark model" and, as expected, we find that these randomization blocks are predictive of treatment. Second, we use the same aggressive algorithm to predict unit-level treatment status using district-division-year randomization blocks plus a collection of additional pre-treatment variables. We call this our "saturated model."[7] If cases are truly randomly assigned to judges, then these additional pre-treatment variables should not provide any additional predictive power above and beyond the district-division-year randomization blocks.

6. Note that each case can have multiple observations in the FJC's IDB since a case can be terminated and reopened multiple times. For each case (identified by a docket number), we only take one of the entries in the FJC's IDB: the last one that occurs before post-appeal actions. Our rationale for this is that the initially assigned judge may cause the case to be reopened repeatedly if, for example, that judge is prone to dismiss cases for minor defects.

7. Essentially, we estimate a propensity score model using a cross-validated machine learning algorithm that more aggressively targets accurate predictions than the commonly used logistic regression.

Our initial tests revealed an imbalance in the cases heard by Republican appointees, suggesting the possibility of non-random assignment. Further inspection revealed that the imbalance was due to a pattern of case assignments to one judge. As the left panel of Figure B.1 illustrates, upon becoming chief judge for the Southern District of New York (NYSD) in 2009, Judge Loretta Preska (a Republican appointee) began hearing a much larger number of civil rights cases. The center panel strongly suggests that these additional cases were not randomly assigned; the rate at which she granted judgment for defendants also increased precipitously and suddenly. These may have been especially strong cases for defendants. The right panel shows that, when applied only to NYSD cases, our test for imbalance detected severe violations of randomized case assignment (visualized with standard ROC curves for the benchmark and saturated models). We thus dropped from our dataset all cases heard by Judge Preska during her tenure as chief judge.

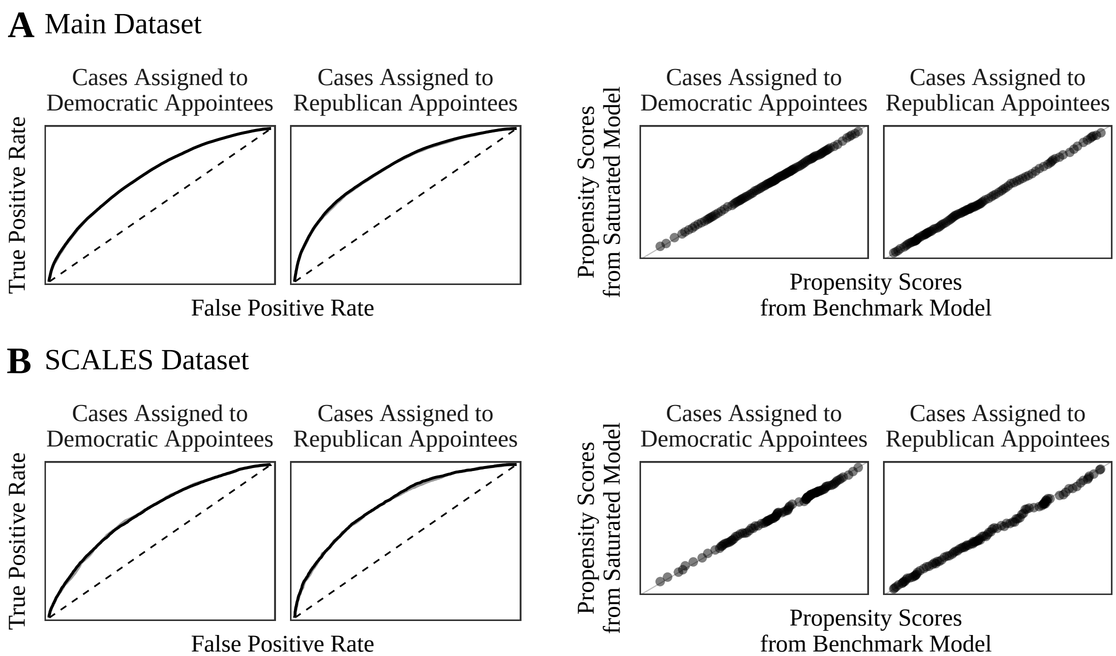
Figure B.1: *The left panel plots the yearly number of cases assigned to each judge in the Southern District of New York (NYSD). Black dots are for Judge Loretta Preska, and gray dots are for all other judges in the court. Triangles indicate which judge is chief judge. The time period when Judge Preska served as chief judge is shaded gray. The middle panel is similar, except it depicts the number of judgments for the defendant issued by each judge in NYSD. Finally, the right panel plots ROC curves for the randomization test described in the main text, when we include cases heard by Judge Preska.*



After dropping these cases, we again performed our machine learning randomization test for cases assigned to Democratic appointees and cases assigned to Republican appointees. We do this for both our main dataset and the SCALES dataset. In all tests, the saturated model does not provide additional predictive power over the benchmark model, supporting the assumption of randomized case assignment. We illustrate this with standard ROC curves plotted in the left half of Figure B.2. In the right half, we present this information in a slightly different way, with eQQ plots comparing the distribution of propensity scores from the benchmark and saturated models. Note that the distributions are nearly identical, again supporting the assumption of randomized case assignment.

All analyses reported in the paper are conducted without cases assigned to Judge Preska during the period she was chief judge. Results don't change if we drop cases heard by all chief judges or drop all cases filed in NYSD during Judge Preska's term as chief judge.

Figure B.2: *We plot the results of our case randomization tests for the main dataset (Panel A) and the SCALES dataset (Panel B). The left plots show ROC curves for the benchmark (gray lines) and saturated models (black lines) described in the main text. (Note: the gray lines are almost completely covered by the black ones.) The right panels are eQQ plots comparing the distributions of propensity scores from the benchmark and saturated models.*



C Additional Analyses and Robustness Checks

Figure C.1: *Each point plots an average treatment effect on defendant wins, 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. For each set of estimates, we present the number of cases in our analysis (top number) and the number of treatment/control judges (bottom number). Note: as in all our analyses, we only provide estimates if there are at least 20 judges in the treatment group. Full results for this plot are available in Table E.6 in Online Appendix E.*

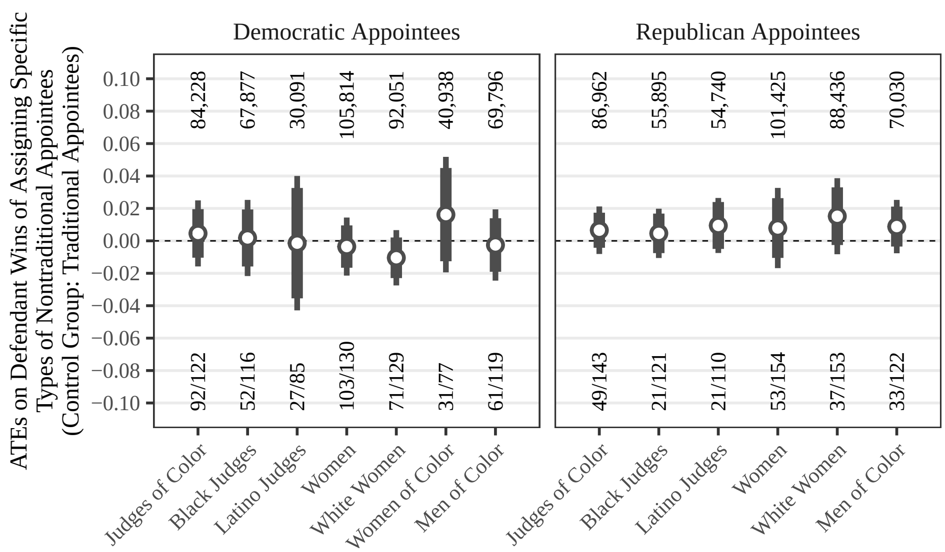


Figure C.2: *Purple circles display average treatment effects of assigning a case to a subgroup of nontraditional appointees (relative to traditional appointees) in the set of cases brought by plaintiffs who share the identity of the treatment judges. Green squares display effects in the set of cases brought by plaintiffs who do not share the identity of the treatment judges. Darker purple/green indicates we used the* `wru` *package to code plaintiff race and lighter purple/green indicates we used the* `predictrace` *package. The 95% confidence intervals are not adjusted for multiple hypothesis testing. Full results for this plot are available in Table E.7 in Online Appendix E.*

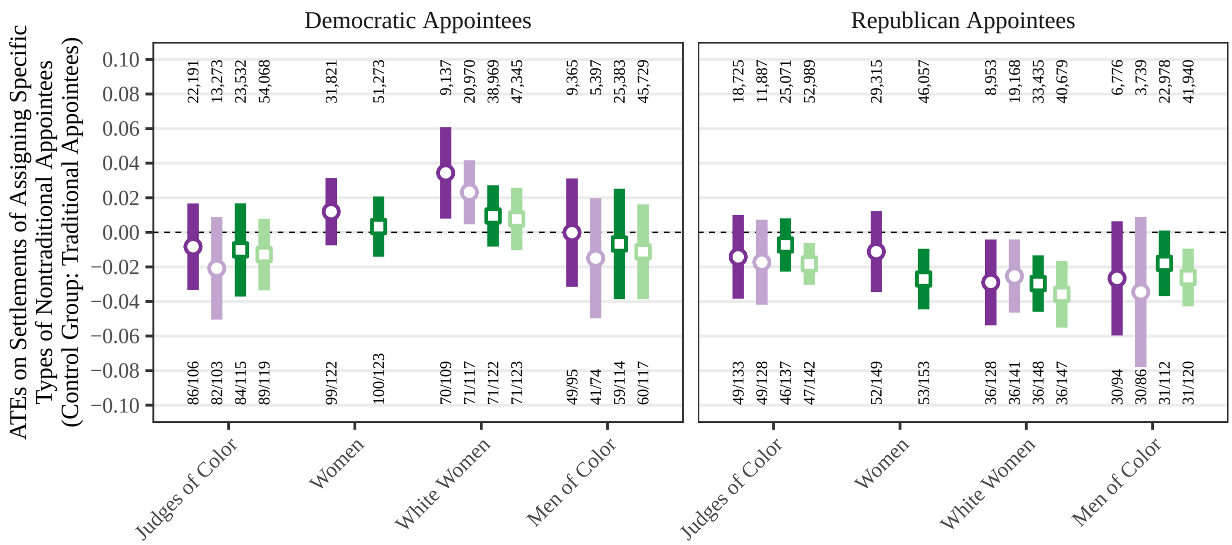
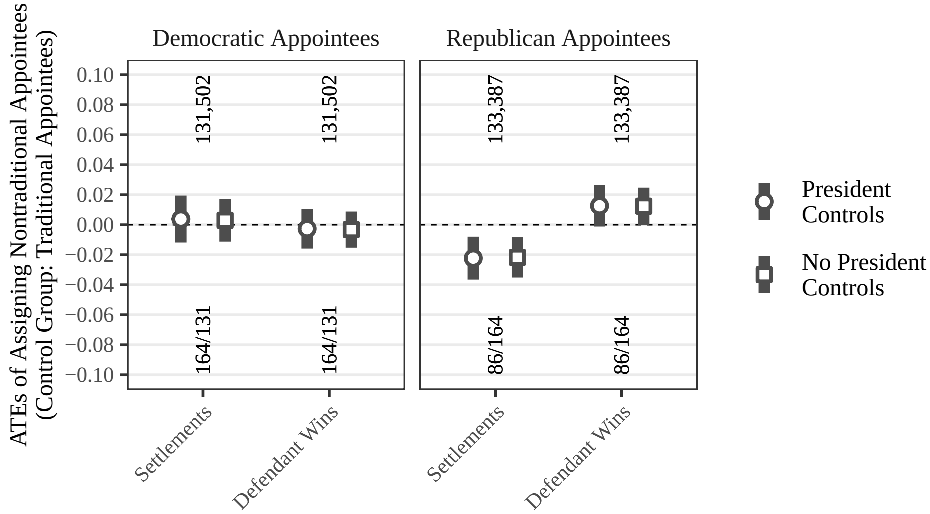


Figure C.3: *Average treatment effects for assignment to nontraditional appointees, with and without controls for appointing president, along with 95% confidence intervals. Full results for this plot are available in Table E.5 in Online Appendix E.*



D Formal Model of Trading Diversity

We analyze a game that resembles a classic agenda-setting model, most prominently articulated by Romer and Rosenthal (1978). Our model presupposes a judicial vacancy and features two players, D and R, who we index by i. The game begins with one player being chosen to be the president (P); the other is the opposition party in the Senate (O). The president proposes a nominee, and the Senate must decide whether to approve the nominee. We abstract away from internal Senate decision-making and simply assume that the key decision-maker is the opposition party.

Sequence The game proceeds as follows:

1. Nature chooses $P \in \{D, R\}$ and b_P, which are publicly revealed.
2. $P \in \{D, R\}$ chooses nominee ideology $x \in \mathbb{R}$ and whether they are from an underrepresented group, $d \in \{0, 1\}$.
3. Nature reveals whether the nominee is qualified, $q \in \{0, 1\}$, where Pr$(q = 0) \equiv \nu < 1/2$.
4. O decides whether to support or oppose the nominee, $s \in \{0, 1\}$.
5. If O supports, the nominee is appointed, otherwise, the players receive default payoffs.[8]
6. Payoffs are realized.

We allow Nature to choose who is president and the benefit that president gets from diversity. This uncertainty plays no role in the players' strategic calculations as the information is publicly revealed. We include this step so that we can talk more clearly about the "likelihood" that the president nominates a nontraditional appointee. For completeness, we assume P is drawn from a binomial distribution with Pr$(P = R) = \rho$ and Pr$(P = D) = 1 - \rho$ and b_P is drawn from some distribution with strictly positive mass on $\mathbb{R}$, and with cdf F_P that depends on the party of the president.

Nominees There is a pool of available nominees for each party: $X \times D = \mathbb{R} \times \{0, 1\}$. A nominee is a pair, (x, d), indicating their ideology and whether they are a nontraditional appointee.

We allow for the possibility that the nominee is discovered to be unqualified during the Senate's review of the nominee's qualifications. Formally, after the president announces the nomination, Nature reveals whether the nominee is qualified, $q \in \{0, 1\}$, where Pr$(q = 0) = \nu < 1/2$. We assume that the nominee is more likely to be qualified than not qualified. This extra step in the game does not fundamentally alter the players' strategic calculations, but it does ensure that there will be rejections in equilibrium.

8. These could be payoffs corresponding to the nominee being rejected, or other kinds of political costs.

Players and payoffs The payoff function for a player $i \in \{D, R\}$ is:

$$u_i = \begin{cases} (b_i - c_i p_i)d - (x - \hat{x}_i)^2 - (1-q)\kappa_i & \text{if nominee is accepted} \\ \overline{u}_i & \text{if nominee is rejected} \end{cases}$$

where:

- $p_i \in \{0, 1\}$ is an indicator variable for whether i is the president/proposer.
- $x \in X = \mathbb{R}$ is the ideology of the nominee, and $d \in D = \{0, 1\}$ indicates whether the nominee is considered a "nontraditional" nominee (see discussion of this terminology in the main text).
- $b_i \in \mathbb{R}$ is the benefit that i gets from a nontraditional appointee relative to a traditional appointee. Note: this allows for the case where $b_i < 0$, indicating a preference for traditional appointees.
- $c_i \geq 0$ is the "search cost" for nominating a nontraditional nominee. Notes: (1) this is only paid if i is the proposer, and (2) this can vary by party.
- $\overline{u}_i \in \mathbb{R}$ is i's default payoff from a nominee being opposed by the opposition party. We interpret this parameter as a measure of the "strength" of the opposition party.
- $\kappa_i > 0$ is the cost associated with an "unqualified" nominee being appointed.

We make the following scope assumptions on the payoffs.

Assumption D.1. Each player strictly prefers a nominee that is at their own ideal point (regardless of diversity concerns). Formally, min$\{0, b_i\} > \overline{u}_i$.

Assumption D.2. Each player strictly prefers a nomination fails if the ideology of the nominee is at the other player's ideal point. Formally, max$\{0, b_i\} - (\hat{x}_j - \hat{x}_i)^2 < \overline{u}_i$.

Assumption D.3. Each player strictly prefers a nomination fails if the nominee is revealed to be an unqualified candidate. Formally, max$\{0, b_i\} + \kappa_i < \overline{u}_i$.

Assumption D.1 ensures that each party's most preferred outcome is a nominee at its ideal point. Assumption D.2 allows us to rule out corner solutions in which a president can nominate a candidate at her own ideal point. This ensures that the president faces a genuine ideological trade-off when she makes a nomination. Assumption D.3 ensures that the opposition always rejects a candidate who is revealed to be unqualified.

We will characterize subgame perfect equilibria, which we find using backward induction. As is standard, we will denote an equilibrium strategy with a star and a generic strategy without a star.

Senate's strategy O supports a nominee (x, d) if and only if

$$b_O d - (x - \hat{x}_O)^2 - (1 - q)\kappa_i \geq \overline{u}_O$$

The $\Leftarrow$ direction is obvious. However, as is standard in agenda-setting models, the $\Rightarrow$ direction will hold in any equilibrium since if O rejects when indifferent, P has no maximizer.

Immediately, by Assumption D.3, $s^* = 0$ if $q = 0$.

Next, consider the case in which $q = 1$. Let $\tilde{x}(d)$ indicate the most ideologically congruent nominee P can get (as a function of d) while satisfying O's constraint, which is implicitly defined by:

$$b_O d - (\tilde{x}(d) - \hat{x}_O)^2 = \overline{u}_O \tag{D.1}$$

Since $b_O d > \overline{u}_O$ (by Assumption D.1),[9] we can solve for the x that induces acceptance:

$$\tilde{x}(d) = \begin{cases} \hat{x}_O + \sqrt{b_O d - \overline{u}_O} & \text{if } \hat{x}_O < \hat{x}_P \\ \hat{x}_O - \sqrt{b_O d - \overline{u}_O} & \text{if } \hat{x}_O > \hat{x}_P \end{cases}$$

Lemma D.1. Given Assumption D.1 and Assumption D.3, $s^*(x, d, q) = 1$ if and only if $q = 1$ and $|x - \hat{x}_O| \leq |\tilde{x}(d) - \hat{x}_O|$.

Proof. In the preceding text. □

President's strategy First, assume that P satisfies O's constraint.

Since O never supports an unqualified nominee ($q = 0$), this induces some uncertainty for P. Then, P's ex ante expected payoff from appointing a nominee (x, d) that satisfies O's constraint is

$$(1 - \nu)[(b_P - c_P)d - (x - \hat{x}_P)^2] + \nu\overline{u}_P$$

By Assumption D.2, O always rejects a nominee with $x = \hat{x}_P$, so P will seek to get the most ideologically congruent judge she can possibly get.

We make the following assumption to simplify the exposition by reducing the number of substantively trivial cases we need to consider. Note that this only has bite for a knife-edge region of the parameter space.

Assumption D.4. When indifferent, the president chooses $d = 1$.

9. Note: if this condition were to fail, then (D.1) has no solution, and O cannot be induced to accept the nominee.

Then, given Assumption D.4, P nominates a nontraditional nominee if and only if:

$$b_P - c_P - (\tilde{x}(1) - \hat{x}_P)^2 \geq -(\tilde{x}(0) - \hat{x}_P)^2$$

(Note: the weakness of this inequality comes from Assumption D.4.) Whether this condition holds depends on the relative weight P places on ideology and diversity. If $b_P > c_P$, then diversity directly increases P's payoff. On the other hand, if $b_P < c_P$, then diversity directly lowers P's payoffs. Rearranging yields:

$$b_P \geq c_P + (\tilde{x}(1) - \hat{x}_P)^2 - (\tilde{x}(0) - \hat{x}_P)^2 \equiv \hat{b}_P \tag{D.2}$$

Let $\hat{b}_P$ be the value of b_P such that the condition binds. Then, if $b_P \geq \hat{b}_P$, then P will nominate a nontraditional appointee. Note $\hat{b}_P$ can be (weakly) negative, so it is possible that the president cannot be induced to nominate a nontraditional nominee since we require $b_P > 0$. When does this happen?

Case 1: Suppose $b_O > 0$. Then, it is straightforward to see that P can get a more ideologically congruent appointment by nominating a nontraditional nominee: $|\tilde{x}(1) - \hat{x}_P| < |\tilde{x}(0) - \hat{x}_P|$. Then, the right-hand side of (D.2) is strictly negative.

Case 1A: Suppose that $b_P - c_P > 0$. Then, P will always nominate a nontraditional nominee, even if the nontraditional nominee is not more ideologically congruent (i.e., if $\tilde{x}(0) \approx \tilde{x}(1)$).

Case 1B: Suppose that $b_P - c_P < 0$. Then, P will nominate a nontraditional nominee if and only if the nontraditional nominee is more ideologically congruent. Moreover, as $b_P - c_P$ declines, P requires a more ideologically congruent nominee in order to be willing to appoint a nontraditional nominee.

Case 2: Suppose $b_O < 0$. Then, it is straightforward to see that P can get a more ideologically congruent appointment by nominating a traditional nominee: $|\tilde{x}(1) - \hat{x}_P| > |\tilde{x}(0) - \hat{x}_P|$. This is because the Senate is biased against nontraditional nominees. Then, the right-hand side of (D.2) is strictly positive. So, at a minimum, P must value diversity in order for the condition to hold. Moreover, to overcome O's opposition (i.e, the relatively larger gap between $\tilde{x}(0)$ and $\tilde{x}(1)$), she must *highly* value diversity in order to be willing to nominate an appointee representing a nontraditional group.

Case 3: Suppose $b_O = 0$. Then, it is straightforward to see that P cannot get a more ideologically congruent appointment by nominating a nontraditional or traditional nominee: $|\tilde{x}(1) - \hat{x}_P| = |\tilde{x}(0) - \hat{x}_P|$. In this case, the Senate gets no positive or negative payoff from diversity. Then, the right-hand side of (D.2) is zero and P is willing to appoint a nontraditional appointee if $b_P > c_P$.

Recall the above analysis proceeded by supposing that P satisfies O's acceptance constraint. We now characterize the conditions under which this occurs. First note that if P satisfies O's constraint, she will select either $(\tilde{x}(1), 1)$ or $(\tilde{x}(0), 0)$. Let:

$$\tilde{U}_P(d) = \begin{cases} -(\tilde{x}(0) - \hat{x}_P)^2 & \text{if } d = 0 \\ (b_P - c_P) - (\tilde{x}(1) - \hat{x}_P)^2 & \text{if } d = 1 \end{cases}$$

Let $\overline{u}_P^{\text{accept}}$ be defined by:

$$\overline{u}_P^{\text{accept}} \equiv \min\{\tilde{U}_P(0), \tilde{U}_P(1)\}$$

Then, it is weakly optimal for P to satisfy O's constraint if $\overline{u}_P^{\text{accept}} \geq \overline{u}_P$, and strictly optimal if the condition holds strictly. In the spirit of classical bargaining models, we opt to focus on cases in which rejection is the worst outcome for both players. So, we make this additional assumption:

Assumption D.5. $\overline{u}_P^{\text{accept}} > \overline{u}_P$.

We can now characterize a unique equilibrium of the model.

Proposition D.1. Given Assumptions D.1–D.5, there is a unique subgame perfect equilibrium of the game that is characterized as follows:

- O supports the nominee ($s^* = 1$) if and only if $q = 1$ and $|x - \hat{x}_O| \leq |\tilde{x}(d) - \hat{x}_O|$.
- P proposes a nominee (x^*, d^*) such that $x^* = \tilde{x}(d)$ and $d^* = 1$ if and only if $b_P \geq \hat{b}_P$.

Proof of Proposition D.1. In the preceding text. □

Note that if Assumption D.5 fails, then it is possible (although not guaranteed) to get rejection on the equilibrium path if there is no nominee P could choose to make her better off than her default payoff. Clearly, this would be a substantively strange situation, as it implies the president would be better off with her nominees being opposed/rejected.

Empirical Implications In the remainder, we characterize several empirical implications of our model, in addition to the one in the main text.

Proposition 1. *In the main text.*

Proof of Proposition 1. Follows directly from comparing (D.1) when setting $d = 1$ and when $d = 0$. □

Let δ^* be the ex ante equilibrium probability that P nominates a nontraditional nominee: $\delta^* \equiv \Pr(b_P > \hat{b}_P) = 1 - F_P(\hat{b}_P)$.

Proposition D.2. As the cost of nominating a nontraditional appointee (c_P) increases, the president is less likely to appoint a nontraditional appointee (δ^* decreases).

Proof of Proposition D.2. First, note that $\hat{b}_P$ (defined in equation D.2) increases as c_P increases. Since F_P is a cdf on a distribution with positive mass on $\mathbb{R}$, it is increasing in its argument. Then, $\delta^* = 1 - F_P(\hat{b}_P)$ is decreasing in $\hat{b}_P$. $\square$

E Regression Tables

On the last pages of this appendix, we present regression tables corresponding to the average treatment effects we plot in the main text and the preceding sections of the Online Appendix.

References

Blevins, Cameron, and Lincoln Mullen. 2015. "Jane, John ... Leslie? A Historical Method for Algorithmic Gender Prediction." *Digital Humanities Quarterly* 9 (3).

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.

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

Imai, Kosuke, and Kabir Khanna. 2016. "Improving Ecological Inference by Predicting Individual Ethnicity from Voter Registration Records." *Political Analysis* 24 (2): 263–272.

Kaplan, Jacob. 2021. *predictrace: Predict the Race and Gender of a Given Name Using Census and Social Security Administration Data.* https://github.com/jacobkap/predictrace/.

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.

Romer, Thomas, and Howard Rosenthal. 1978. "Political Resource Allocation, Controlled Agendas and the Status Quo." *Public Choice* 33 (4): 27–43.

Table E.1: *Results from the models used to estimate the effects in Figure 4*

	Democratic Appointees		Republican Appointees		All Appointees	
	(1)	(2)	(3)	(4)	(5)	(6)
Nontraditional Appointees	0.003	–0.003	–0.022**	0.012*	—	—
	(0.007)	(0.006)	(0.007)	(0.006)		
Republican Appointees	—	—	—	—	–0.016**	0.009
					(0.005)	(0.005)
Outcome	Stlmt.	Deft. Wins	Stlmt.	Deft. Wins	Stlmt.	Deft. Wins
Cases	131,502	131,502	133,387	133,387	291,243	291,243
Treatment Judges	164	164	86	86	264	264
Control Judges	131	131	164	164	304	304
Randomization Blocks	✓	✓	✓	✓	✓	✓
Min. Units/Trt. Arm	5	5	5	5	5	5
Appt. Pres. Controls						

Notes: All models include district-division-year fixed effects, which we refer to as our "randomization blocks." We also use an adjustment proposed by Lin (2013) for all control variables (see main text). Standard errors are clustered by judge. Statistical significance is indicated by stars: * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

Table E.2: *Results from the models used to estimate the effects in Figure 5*

	Democratic Appointees		Republican Appointees	
	(1)	(2)	(3)	(4)
Nontraditional Appointees	–0.01	–0.002	–0.034**	0.022
	(0.011)	(0.011)	(0.011)	(0.012)
Outcome	Stlmt.	Deft. Wins	Stlmt.	Deft. Wins
Cases	26,033	26,033	17,411	17,411
Treatment Judges	215	215	104	104
Control Judges	174	174	168	168
Randomization Blocks	✓	✓	✓	✓
Min. Units/Trt. Arm	5	5	5	5
Appt. Pres. Controls				

Notes: All models include district-division-year fixed effects, which we refer to as our "randomization blocks." We also use an adjustment proposed by Lin (2013) for all control variables (see main text). Standard errors are clustered by judge. Statistical significance is indicated by stars: * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

Table E.3: *Results from the models used to estimate the effects in Figure 6*

	Democratic Appointees							Republican Appointees					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Judges of Color	–0.012	—	—	—	—	—	—	–0.012	—	—	—	—	—
	(0.01)							(0.006)					
Black Judges	—	–0.017	—	—	—	—	—	—	–0.009	—	—	—	—
		(0.011)							(0.007)				
Latino Judges	—	—	–0.003	—	—	—	—	—	—	–0.015	—	—	—
			(0.023)							(0.006)			
Women	—	—	—	0.006	—	—	—	—	—	—	–0.015	—	—
				(0.008)							(0.008)		
White Women	—	—	—	—	0.015	—	—	—	—	—	—	–0.028*	—
					(0.007)							(0.008)	
Women of Color	—	—	—	—	—	–0.025	—	—	—	—	—	—	—
						(0.017)							
Men of Color	—	—	—	—	—	—	–0.004	—	—	—	—	—	–0.022*
							(0.011)						(0.007)
Outcome	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.
Cases	84,228	67,877	30,091	105,814	92,051	40,938	69,796	86,962	55,895	54,740	101,425	88,436	70,030
Treatment Judges	92	52	27	103	71	31	61	49	21	21	53	37	33
Control Judges	122	116	85	130	129	77	119	143	121	110	154	153	122
Randomization Blocks	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Min. Units/Trt. Arm	5	5	5	5	5	5	5	5	5	5	5	5	5
Appt. Pres. Controls													

Notes: All models include district-division-year fixed effects, which we refer to as our "randomization blocks." We also use an adjustment proposed by Lin (2013) for all control variables (see main text). Standard errors are clustered by judge. To adjust for multiple comparisons, we estimated the number of independent tests that we conducted and applied a Bonferroni correction (see the text). Statistical significance (adjusted for multiple comparisons) is indicated by stars: * $p < 0.0063$, ** $p < 0.0013$ and *** $p < 0.0001$.

Table E.4: *Results from the models used to estimate the effects in Figure 7*

	Democratic Appointees				Republican Appointees			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Judges of Color	–0.01	–0.008	—	—	–0.007	–0.014	—	—
	(0.013)	(0.012)			(0.008)	(0.012)		
Women	—	—	0.003	0.012	—	—	–0.027**	–0.011
			(0.009)	(0.01)			(0.009)	(0.011)
Outcome	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.
Cases	23,532	22,191	51,273	31,821	25,071	18,725	46,057	29,315
Treatment Judges	84	86	100	99	46	49	53	52
Control Judges	115	106	123	122	137	133	153	149
Randomization Blocks	✓	✓	✓	✓	✓	✓	✓	✓
Min. Units/Trt. Arm	5	5	5	5	5	5	5	5
Appt. Pres. Controls								
Pltf. Shares Identity	No	Yes	No	Yes	No	Yes	No	Yes
Race Coding	w	w			w	w		

Notes: All models include district-division-year fixed effects, which we refer to as our "randomization blocks." We also use an adjustment proposed by Lin (2013) for all control variables (see main text). Standard errors are clustered by judge. Statistical significance is indicated by stars: * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. "Pltf. Shares Identity" indicates whether the analysis is on subset of cases in which the plaintiff(s) share the identity of the treatment judges, or not. "Race Coding" indicates which package was used to predict plaintiff race: "w" is `wru` and "p" is `predict race`. For plaintiff gender we use the `gender` package.

Table E.5: *Results from the models used to estimate the effects in Figure C.3*

	Democratic Appointees				Republican Appointees			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Nontraditional Appointees	0.003	0.004	–0.003	–0.003	–0.022**	–0.022**	0.012*	0.013
	(0.007)	(0.008)	(0.006)	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)
Outcome	Stlmt.	Stlmt.	Deft. Wins	Deft. Wins	Stlmt.	Stlmt.	Deft. Wins	Deft. Wins
Cases	131,502	131,502	131,502	131,502	133,387	133,387	133,387	133,387
Treatment Judges	164	164	164	164	86	86	86	86
Control Judges	131	131	131	131	164	164	164	164
Randomization Blocks	✓	✓	✓	✓	✓	✓	✓	✓
Min. Units/Trt. Arm	5	5	5	5	5	5	5	5
Appt. Pres. Controls		✓		✓		✓		✓

Notes: For the appointing president variable, the excluded categories are Reagan and Clinton. All models include district-division-year fixed effects, which we refer to as our "randomization blocks." We also use an adjustment proposed by Lin (2013) for all control variables (see main text). Standard errors are clustered by judge. Statistical significance is indicated by stars: * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

Table E.6: *Results from the models used to estimate the effects in Figure C.1*

	Democratic Appointees							Republican Appointees					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Judges of Color	0.005 (0.007)	—	—	—	—	—	—	0.007 (0.005)	—	—	—	—	—
Black Judges	—	0.002 (0.009)	—	—	—	—	—	—	0.005 (0.006)	—	—	—	—
Latino Judges	—	—	–0.001 (0.015)	—	—	—	—	—	—	0.009 (0.006)	—	—	—
Women	—	—	—	–0.004 (0.007)	—	—	—	—	—	—	0.008 (0.009)	—	—
White Women	—	—	—	—	–0.01 (0.006)	—	—	—	—	—	—	0.015 (0.009)	—
Women of Color	—	—	—	—	—	0.016 (0.013)	—	—	—	—	—	—	—
Men of Color	—	—	—	—	—	—	–0.003 (0.008)	—	—	—	—	—	0.009 (0.006)
Outcome	Deft. Wins	Deft. Wins	Deft. Wins	Deft. Wins	Deft. Wins	Deft. Wins	Deft. Wins	Deft. Wins	Deft. Wins	Deft. Wins	Deft. Wins	Deft. Wins	Deft. Wins
Cases	84,228	67,877	30,091	105,814	92,051	40,938	69,796	86,962	55,895	54,740	101,425	88,436	70,030
Treatment Judges	92	52	27	103	71	31	61	49	21	21	53	37	33
Control Judges	122	116	85	130	129	77	119	143	121	110	154	153	122
Randomization Blocks	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Min. Units/Trt. Arm	5	5	5	5	5	5	5	5	5	5	5	5	5
Appt. Pres. Controls													

Notes: All models include district-division-year fixed effects, which we refer to as our "randomization blocks." We also use an adjustment proposed by Lin (2013) for all control variables (see main text). Standard errors are clustered by judge. To adjust for multiple comparisons, we estimated the number of independent tests that we conducted and applied a Bonferroni correction (see the text). Statistical significance (adjusted for multiple comparisons) is indicated by stars: * $p < 0.0063$, ** $p < 0.0013$ and *** $p < 0.0001$.

Table E.7: *Results from the models used to estimate the effects in Figure C.2*

	Democratic Appointees													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Judges of Color	–0.01	–0.008	–0.013	–0.021	—	—	—	—	—	—	—	—	—	—
	(0.013)	(0.012)	(0.01)	(0.015)										
Women	—	—	—	—	0.003	0.012	—	—	—	—	—	—	—	—
					(0.009)	(0.01)								
White Women	—	—	—	—	—	—	0.009	0.034*	0.008	0.023*	—	—	—	—
							(0.009)	(0.013)	(0.009)	(0.009)				
Men of Color	—	—	—	—	—	—	—	—	—	—	–0.007	0	–0.011	–0.015
											(0.015)	(0.015)	(0.013)	(0.017)
Outcome	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.
Cases	23,532	22,191	54,068	13,273	51,273	31,821	38,969	9,137	47,345	20,970	25,383	9,365	45,729	5,397
Treatment Judges	84	86	89	82	100	99	71	70	71	71	59	49	60	41
Control Judges	115	106	119	103	123	122	122	109	123	117	114	95	117	74
Randomization Blocks	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Min. Units/Trt. Arm	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Appt. Pres. Controls														
Pltf. Shares Identity	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Race Coding	w	w	p	p			w	w	p	p	w	w	p	p

	Republican Appointees													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Judges of Color	–0.007	–0.014	–0.018**	–0.017	—	—	—	—	—	—	—	—	—	—
	(0.008)	(0.012)	(0.006)	(0.012)										
Women	—	—	—	—	–0.027**	–0.011	—	—	—	—	—	—	—	—
					(0.009)	(0.011)								
White Women	—	—	—	—	—	—	–0.03***	–0.029*	–0.036***	–0.025*	—	—	—	—
							(0.008)	(0.012)	(0.009)	(0.01)				
Men of Color	—	—	—	—	—	—	—	—	—	—	–0.018	–0.027	–0.026**	–0.035
											(0.009)	(0.016)	(0.008)	(0.021)
Outcome	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.	Stlmt.
Cases	25,071	18,725	52,989	11,887	46,057	29,315	33,435	8,953	40,679	19,168	22,978	6,776	41,940	3,739
Treatment Judges	46	49	47	49	53	52	36	36	36	36	31	30	31	30
Control Judges	137	133	142	128	153	149	148	128	147	141	112	94	120	86
Randomization Blocks	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Min. Units/Trt. Arm	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Appt. Pres. Controls														
Pltf. Shares Identity	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Race Coding	w	w	p	p			w	w	p	p	w	w	p	p

Notes: All models include district-division-year fixed effects, which we refer to as our "randomization blocks." We also use an adjustment proposed by Lin (2013) for all control variables (see main text). Standard errors are clustered by judge. Statistical significance is indicated by stars: * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. "Pltf. Shares Identity" indicates whether the analysis is on subset of cases in which the plaintiff(s) share the identity of the treatment judges, or not. "Race Coding" indicates which package was used to predict plaintiff race: "w" is `wru` and "p" is `predict race`. For plaintiff gender we use the `gender` package.