DISCUSSION PAPER SERIES

DP14370

PROFESSIONAL INTERACTIONS AND HIRING DECISIONS: EVIDENCE FROM THE FEDERAL JUDICIARY

Marco Battaglini, Jorgen Harris and Eleonora Patacchini

LABOUR ECONOMICS

ISSN 0265-8003

PROFESSIONAL INTERACTIONS AND HIRING DECISIONS: EVIDENCE FROM THE FEDERAL JUDICIARY

Marco Battaglini, Jorgen Harris and Eleonora Patacchini

Discussion Paper DP14370
Published 31 January 2020
Submitted 29 January 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

PROFESSIONAL INTERACTIONS AND HIRING DECISIONS: EVIDENCE FROM THE FEDERAL JUDICIARY

Abstract

We examine the effect of hearing cases alongside female judicial colleagues on the probability that a federal judge hires a female law clerk. Federal judges are assigned to cases and to judicial panels at random and have few limitations on their choices of law clerks: these two features make the federal court system a unique environment in which to study the effect of professional interactions and beliefs in organizations. We constructed a unique dataset by aggregating federal case records from 2007-2017 to collect information on federal judicial panels, and by merging this data with judicial hiring information from the Judicial Yellow Book, a directory of federal judges and clerks. We find that a one standard deviation increase in the fraction of co-panelists who are female increases a judge's likelihood of hiring a female clerk by 4 percentage points. This finding suggests that increases in the diversity of the upper rungs of a profession can shift attitudes in a way that creates opportunities at the entry level of a profession.

JEL Classification: J16, J71, J82

Keywords: Economics of gender, discrimination, Labor Force Composition

Marco Battaglini - battaglini@cornell.edu
Cornell University and CEPR

Jorgen Harris - jorgenharris@oxy.edu
Occidental College

Eleonora Patacchini - eleonora.patacchini14@gmail.com
Cornell University and CEPR

Professional Interactions and Hiring Decisions: Evidence from the Federal Judiciary[1]

Marco Battaglini
Cornell University

Jorgen M. Harris
Occidental College

Eleonora Patacchini
Cornell University

Abstract

We examine the effect of hearing cases alongside female judicial colleagues on the probability that a federal judge hires a female law clerk. Federal judges are assigned to cases and to judicial panels at random and have few limitations on their choices of law clerks: these two features make the federal court system a unique environment in which to study the effect of professional interactions and beliefs in organizations. We constructed a unique dataset by aggregating federal case records from 2007-2017 to collect information on federal judicial panels, and by merging this data with judicial hiring information from the *Judicial Yellow Book*, a directory of federal judges and clerks. We find that a one standard deviation increase in the fraction of co-panelists who are female increases a judge's likelihood of hiring a female clerk by 4 percentage points. This finding suggests that increases in the diversity of the upper rungs of a profession can shift attitudes in a way that creates opportunities at the entry level of a profession.

JEL Codes: J16, J82, J71

Keywords: Economics of Gender, Discrimination, Labor Force Composition

[1] We thank Naci Mocan for helpful comments. We also thank Sam Markiewitz, Becky Cardinali, Cenjia Lu, and Ignasi Merediz for excellent research assistance. Harris and Patacchini acknowledge financial support from the Institute for Social Sciences at Cornell University.

1. Introduction

Prejudiced and discriminatory beliefs are widespread in human relationships, have profoundly negative consequences both for members of out-groups and for society as a whole, and may lead to inefficient economic decisions. A long-standing literature in sociology documents how these beliefs are formed, when they are maintained, and how they can be disrupted. These topics have attracted growing attention among economists. Several theoretical papers have developed models explaining the perpetuation of different beliefs between groups under different hypotheses, including the scenario when one group's beliefs are incorrect. Still the empirical literature has struggled to formally document causal effects of such beliefs on economic outcomes.

In this paper, we exploit a unique institutional feature of the Federal appellate court system to present clear causal evidence on the effect of exposure to out-groups for individual hiring decisions. Appellate court cases are typically heard by panels of three judges, randomly selected from a pool of appellate court justices and district court justices. Because appellate judges do not choose the cases that come before their courts or the colleagues with whom they hear these cases, their likelihood of working with female colleagues on cases is effectively random. At the same time, appellate judges are broadly unconstrained in their decision of who to hire as a court clerk, a highly prestigious position typically filled by graduates of top law programs. As a result, changes in the likelihood of hiring a female clerk likely reflect changes in a judge's assessment of the likely ability of a female junior colleague. We can thus exploit exposure to female colleagues in the appellate panels as exogenous shocks to a judge's attitude in order to assess its causal impact on the hiring decision of court clerks. An important and unique feature of this data is that the non-voluntary interactions concern high-stakes, sustained, professional interactions between peers at the elite of their occupations.

We find substantial positive effects of professional interactions with female judges on the likelihood of hiring female clerks. In particular, we find that a one standard deviation increase in the fraction of cases a judge hears with female colleagues increases the likelihood of hiring at least one female clerk in the next year by 4 percentage points. Our finding is robust to a wide range of tests, including placebo regressions in which we match each judge to the most similar judge within the court and regress the judge's exposure to female colleagues on the hiring decision of the match.

We next study whether the effect of the interaction with female colleagues depends on the characteristics of the influenced judge and the characteristics of the interacting female colleagues. We observe seven factors that may affect the susceptibility of judges to interactions: the judge's gender, age, experience, political party, status, quality, and the fraction of a judge's staff currently composed of women. For all these factors, the sign of the estimated effect is consistent with the theory that judges update their beliefs following the interactions. We find suggestive evidence that these effects are larger for male judges, for judges with few women currently on their staff, and for less experienced judges.

Our work relates to three strands of literature. The first examines the effect of attitudes toward gender roles on labor market outcomes. This literature highlights the role played by such attitudes as important determinants of the gender wage gap and labor participation. Guiso, Sapienza and Zingales (2003) and Algan and Cahuc (2006) use international individual value surveys to show that religious beliefs are associated with a less positive attitude toward working women. Using data from the World Value surveys on OECD countries, Fortin (2005) shows that anti-egalitarian views are negatively correlated with female employment rates and positively correlated with gender pay gaps. By relying on cross-sectional evidence at the country level, this pioneering literature has however not fully addressed the question of the identification of a causal relationship between gender roles attitudes and labor conditions for women.

The second literature to which our work is related is that on the effects of integration of groups that differ on ethnic, religious or gender dimensions. This work has directly addressed the issue of causality by evaluating the effect of quasi-experimental changes in integration on social and cultural attitudes. See among others, Van Laar et al. (2005), Boisjoly et al. (2006), Corno et al. (2019), Mark and Harris (2012), Carrell et al. (2018), Dahl et al. (2018) and Paluck et al. (2017). The effects of integration are measured in these works with self-reported attitudes or the Implicit Association Test.

The third and most closely related literature to our work studies the effect of integration and exposure to out groups on economically relevant choices, such as hiring decisions or gender and racial wage gaps. Beaman et al. (2009) exploit random assignments of gender quotas for leadership positions in Indian village councils to show that exposure to women in leadership positions affects gender attitudes and that, in the long term, these changes translate in electoral wins for women.[2] Washington (2008) and Glynn and Sen (2015) study the effect of having daughters on the votes of lawmakers and judges in the U.S. court of Appeals. Reuben, Sapienza and Zingales (2014) find in a laboratory experiment that men and women greatly increase their likelihood of hiring a female employee for a simple arithmetic task when they are able to see employees' scores on the task.

Relatedly, recent works examine whether discrimination depends on the gender of evaluators by estimating the gender gap in the recruitment of professors in Italy and Spain where evaluators are randomly selected into committees. Using data on all competitions in Italy and

---

[2] See also Pande and Ford (2011) for a survey of the equity and efficiency impacts of gender quotas for political positions and corporate board membership.

Spain, Bagues et al. (2017) show that committees with more women reduce gender discrimination against women in competitions for full professor positions, but do not affect competitions for associate professor positions. Bagues and Esteve-Volart (2010) find that female candidates are less likely to be hired in the Spanish Judiciary when the randomly assigned selection committee has a higher percentage of female evaluators.

2. Institutional Setting

2.1. Federal Appellate Courts

The US federal court system has three tiers: district courts, appellate courts, and the Supreme Court. All cases are initially heard in district courts, where evidence is presented, parties appear, and an initial ruling is made. Parties bringing a case in federal court are entitled to appeal decisions to appellate courts, which review the legal reasoning used in district courts.

Federal appellate courts are organized into circuits, which are primarily organized geographically. Each appellate court hears appeals generated by district courts within their jurisdiction. As a consequence, federal judges must maintain expertise in a wide variety of legal areas, and might learn different information about the ability of their colleagues in each case that they hear.

Most appellate cases are heard in panels of three judges, with some cases heard by larger panels. While each circuit intends to assign judges to cases at random, the assignment mechanism varies by panel, including the use of computer programs and the drawing of lots (Levy 2019). Some circuit courts, such as the Fifth Circuit, choose panels in a manner that avoids having any judge serve too often with any other judge (Levy 2019), limiting the variance in exposure to other judges in our sample.

Appellate panels consist of regular appellate judges, senior appellate judges (appellate judges who work part time), and visiting judges—typically retired federal judges or district judges serving on a district court that is subsidiary to the appellate court. Though the use of visiting judges varies from court to court, circuits typically ensure that at least two regular appellate judges hear each case (Levy 2019). In addition, visiting judges are often restricted in the cases they hear, for instance only hearing civil cases (Levy 2019), and may not be available year-round. As a result, the exposure of visiting judges to colleagues may differ systematically from the exposure of appellate and senior appellate judges. We therefore restrict our sample to only include appellate and senior appellate judges.

Appellate court decisions create precedent for lower courts. As a result, courts publish their opinions in legal registers, making them available for reference and citation. However, because all litigants are entitled to appellate court review, many cases break no legal ground and are not useful as precedent. As a result, courts only publish cases that, in the judgment of the court, include legal reasoning that is useful for citation. In 2017, fewer than 12% of appellate cases were published. Our analysis includes only these published cases, both because only published cases generate data and because these cases provide judges with more information about the legal reasoning and competence of their colleagues than do routine cases.

2.1.1. Appointment of Judges and Clerks

Appellate judges are appointed to federal courts by receiving a nomination from the President and confirmation from the Senate. Once appointed, judges have lifetime tenure on the court, barring serious misconduct. A majority of appellate court judges have prior judicial experience and 85% have prior experience practicing law (McMillon 2014), and the large majority of nominees have been confirmed under all recent presidents (Gramlich 2018).

Most judges in our sample are over 60 years old and have held their current positions for more than fifteen years. As a result, most federal judges began their careers when women were significantly less represented in the legal profession than they are today. While half of new lawyers are female today, fewer than 5% were female prior to 1968, and only 36% were female in 1981, the year that the median judge in our sample finished law school (American Bar Association 2013). As a result, federal judges make hiring decisions in a labor market where women are significantly more numerous and successful than they were when the judges first entered the legal profession and the judiciary.

Judges have a budget for a staff of law clerks and administrative assistants. Typically hired directly from law school, law clerks generally serve one-year or two-year terms and assist judges with legal research and decision writing (Posner et. al. 2001). Appellate court clerkships are prestigious and competitive positions, and often lead to positions at top law firms, government agencies and the judiciary (Rhinehart 1994).

Despite efforts to delay hiring of clerks to the beginning of their third years, substantial numbers of law clerks are hired or recruited as early as the first semester of their second year of law school (Avery et. al. 2007). As a result, there is as much as a two-year gap between the decision to hire a clerk and the clerk's start date. Given this, we examine the effect of interactions with co-panelists in year t on the gender of clerks who start work in year $t+3$, on the assumption that clerks will start up to two years after their date of hire.

Due to the structure of the appellate clerk market, Judges have wide latitude to choose the candidate clerk that best matches their preferences (Avery et. al. 2007). Because approximately half of all third-year law students apply to clerkship positions, appellate court judges typically receive thousands of applications (NALP 2019). Clerkships are typically offered with very short decision windows (in some cases, as brief as 10 minutes), so most law students take the first clerkship offered.

3. Data description

3.1. Primary Data

We pool data from several sources. Judge and clerk information is collected from the *Judicial Yellow Book* published by Leadership Directories Inc (Leadership Directories 2007-2017). Intended as a resource for attorneys presenting cases in state and federal court, the *Judicial Yellow Book* contains the names and backgrounds of judges serving at all levels of the federal court system, as well as the names of each judge's clerks. We purchased archived copies of the *Judicial Yellow Book* from Leadership Directories Inc. for the years 2007-2017, in the form of pdf pre-publication masters. We use these data to determine the characteristics of judges and the gender of the clerks hired by each judge in each year.

To determine which judges sat on panels together in each year, we scraped information from the online court records aggregator leagle.com (2019). Leagle stores and categorizes the decisions handed published by the federal court system. The library is comprehensive and contains over 5 million published cases since 1950. We pool information on the universe of cases heard between 2007 and 2017, in total 50,813 cases. For each case, Leagle provides the text of the court's decision, exactly as it appears in published court documents. Each document contains headers with the case's docket number and name, the date(s) the case was heard and decided, the court where the case was heard, the attorneys for the appellant and appellee, and most importantly for our purposes the names of the judges who heard that case. We use these records to identify the judges serving on the appellate panel for a given case.

We combine these datasets by matching judges appearing in cases from each circuit court in each year in the case records data with judges listed in the *Judicial Yellow Book* in that circuit court or a subsidiary district court in that year. Judges from subsidiary district courts are included because judges from district courts are invited to serve on appellate panels (28 U.S. Code § 292). 66% of potential judge names identified in the case records are also found in the *Judicial Yellow Book*. The remaining 34% of potential names consist of names of attorneys or parties incorrectly categorized as judges or, in fewer cases, judges visiting from other circuits and retired judges hearing cases as senior judges. These 34% of names are dropped from the analysis. Among appellate court judges in the *Judicial Yellow Book*, 85% appear on at least one case record in the year that they are listed. The majority of judges who do not appear in any published cases are senior judges, and thus have discretion to hear few or no cases in a year. These judges may have only heard unpublished cases, may have taken sick leave, or may be recorded inconsistently in the two data sources. Among district court judges in the *Judicial Yellow Book*, 12% appear on at least one case record, consistent with a significant minority of district judges hearing appellate cases in any particular year. We eliminate judge names that do not match across these two data sources. In total, we identify 298 appellate judges and 589 district judges who served on an appellate court panel that produced a published opinion at least once between 2007-2017.[3] In our final database of 50,484 cases, 70% have three recognized judges, 20% have two judges, 6% have one judge, and 4% have more than three judges. Cases can have fewer than three recognized judges if a member

[3] Judges are identified in court documents by surname only. For nineteen surnames, multiple judges served simultaneously within a circuit (circuit court judges and district judges in subsidiary districts). For twelve of those surnames, the judges were of different genders. In these cases, interactions with a judge of these surnames was counted based on the "expected" gender of the judge. Because appellate judges hear, on average, 32 published cases per year, and district judges hear, on average, 0.5 published cases per year, we take the average gender of judges with each surname, assigning a weight of 32 to appellate judges and a weight of 0.5 to district judges.

of the judicial panel is a visiting judge who is not from a subsidiary district court or if names are recorded inconsistently. Cases can have more than three judges if they are heard en blanc (before all judges on an appellate court) or if a judge was replaced during the progress of the case due to illness or other circumstances.

We ensure the validity of our data by comparing it to official court statistics (presented in online appendix table A.1). First, we compare the count of published cases in our data for each court with the count of published cases reported in annual Judicial Business Tables B.12 from November 2006 to November 2017 (US Courts 2018). We recover between 89% and 98% of published cases for all circuit courts other than the third circuit. We recover only 69% of cases in the Third Circuit. Small deviations above and below the published numbers may be a consequence of January vs November date cutoffs, but records from the Third Circuit are likely incomplete. We also compare the number of judges appearing in the judicial yellow book and hearing cases in each year to the number of judges appearing in the Federal Judicial Center database (Federal Judicial Center 2019) in each year. There are an average of 268 appellate judges appearing in each year of our data, compared to 280 appellate judges in each year of the federal judicial center data. The discrepancy between our data and the Federal Judicial Center data primarily reflects the fact that some judges with senior status are not included in the *Judicial Yellow Book* data. In particular, several judges in the Fourth Circuit are listed in the Judicial Center database as senior appellate judges who had never served as regular appellate judges—none of these judges appear in the *Judicial Yellow Book* data. Likewise, judges who attained senior status prior to 1995 only occasionally appear in the *Judicial Yellow Book* data. Because these judges do not appear in our case records data, we believe that these judges have maintained senior status but are not actively hearing cases. We also incorporate a rating of the conservatism of the president who appointed each judge using the DW Nominate algorithm (Epstein et. al. 2007). We describe the construction of these datasets in greater detail in appendix A.

3.2. Sample Selection

We construct a panel dataset where an observation consists of an appellate court judge in a particular year. As mentioned in Section 2.1, we identify 365 distinct appellate judges in the *Judicial Yellow Book* data, of whom 215 hear at least one appellate case in at least one year of the data (the rest consist of inactive senior judges). If we had records for all eight years for each of these 298 judges, our sample would include 2384 observations. In reality, 20% of judges start after 2007, and another 15% retire before 2014. Of those who started prior to 2007 and continued in their positions until 2014, 5% heard no cases during at least one year of their service. As a result, only 63% of the judges in our sample appear in each year, and we only observe 2158 judge-years of interaction on federal appellate court. Furthermore, 83 judges who hear cases in at least one year hired no appellate clerks during the sample period, and only 8% of the remaining 215 judges hire a clerk every year. As a result, our final sample includes 1074 observations, at the judge by year level, from 215 judges over eight years. As shown in online appendix Table A.2, our sample consists of judges in years where the judge was on at least one panel with a published case, hired at least one clerk in the following year, and is not missing any primary covariates. We also include regressions that control for the current gender composition of a judge's law clerks—this covariate is missing when a judge has no law clerks on staff, resulting in missing values for 87 observations, primarily in the first year of a judge's tenure. As shown in online appendix Table A.3, 26% of observations come from female judges. Judges hire an average of 2.9 law clerks per year, and hire at least one female law clerk in 70% of years in which they make a hiring decision. Overall, 42% of clerks are female.

4. Empirical model and identification strategy

Our empirical strategy takes advantage of the random assignment of judges to panels to regress a measure of interaction with female judges in a given year on the likelihood of hiring female clerks in the following year. Because assignment of appellate judges to cases is random conditional on circuit and year, we control for fixed effects at the circuit by year level. Variation in the female fraction of co-panelists, conditional on court and year, is due entirely to the random assignment of judges to cases, and to the determination of panels that a case is worthy of publication.

Because fewer than 25% of judges hire more than one female clerk per year, we measure propensity to hire female clerks via an indicator of whether at least one female clerk was hired in the year following a judge's exposure. We estimate the effect of exposure to female colleagues on the likelihood of hiring a female clerk using the following regression equation:

$$Hire_{j,c,t+3} = \beta Inf_{j,c,t} + \delta X_{j,c,t} + \theta_{c,t} + \varepsilon_{j,c,t} \qquad (1)$$

where $Hire_{j,c,t+3}$ is a binary indicator of whether judge j, in court c, hired at least one female clerk who started in year $t + 3$; We adopt a variable lagged at $t + 3$ for the hires because, as described in Section 2.1.1, clerks are hired up to 2 years before their actual employment starts. $Inf_{j,c,t}$ is exposure to female judges, $X_{j,c,t}$ is a set of judge characteristics, and $\theta_{c,t}$ is a set of court by year fixed effects.

We measure exposure to female judges $Inf_{j,c,t}$ as the fraction of co-panelists on the cases heard by judge j in year t that are female. There are two noteworthy characteristics of this measure. First, because we calculate the fraction of co-panelists who are female, our measure of exposure to female colleagues does not depend on the total volume of cases heard by a judge in a particular year. This decision reflects two assumptions: that full-time judges with few cases are likely to have more time-consuming cases than those with many cases, and that the salience of individual cases is likely greater when a judge has heard fewer cases in a year. Second, this measure does not distinguish between interactions with a single female colleague on many cases and interactions with multiple female colleagues, each on an individual case. We selected this measure based on a learning model in which each case heard with a co-panelist reveals a small amount of information about that co-panelist's ability, which then informs the likely distribution of legal professionals with the co-panelist's gender. If there is substantial uncertainty about the ability of each judge, repeated interaction with one judge will have similar information content to individual interactions with multiple judges.

In order to demonstrate the feasibility of this approach, we estimate the variation in our main dependent and independent variables that is not accounted for by court by year variation. As shown in Table A.4, very little variation in either the hiring decisions of judges or in the exposure of judges to female colleagues is explained by differences between courts and years. Likewise, observed judge characteristics do not explain a significant amount of variation in either judge hiring decisions or exposure to female judge.

4.1 Evidence in support of identification strategy

The key identifying assumption in this paper is that variation in the gender composition of co-panelists within a particular circuit and year is unrelated to a judge's preference for female clerks and to a judge's available labor pool. This assumption is justified by the assertion, common to all appellate circuits, that judges are randomly assigned to cases (Stearns and Abramowicz 2005). While violation of pure randomness is inevitable, violations of random assignment are small, unlikely to be sustained over a year, seen only in a few courts, and unlikely to be related to judge's preferences or labor pools. In particular, Chilton and Levy (2015) find that due to scheduling conflicts and similar concerns, the assignment of judges to appellate panels deviates from random assignment in several courts. As a consequence, the distribution of Republican appointees across cases differs slightly from what would be expected by chance in the Second, Sixth and DC Circuits, and more substantially in the Ninth Circuit. However, the likelihood that a Republican will serve with another Republican differs from chance by less than a percentage point in all circuits but the Second and Ninth, in which it differs from chance by less than two percentage points. In addition, as shown in Table 1, judges appointed by Republican presidents are as likely to serve on panels with female judges as are judges appointed by Democratic presidents, indicating that any non-randomness in the assignment of judges to panels on the basis of political party does not affect the likelihood of serving on panels with female colleagues. Finally, as shown in Table 2, the inclusion of controls for judge characteristics, including party, does not weaken the measured effect of exposure to female colleagues on hiring decisions. Levy (2017) examines a broader range of potentially non-random scheduling decisions made by the chief justice's office of each appellate circuit. Levy finds, for instance, that one circuit had a tradition of ensuring that judges have the opportunity to be the presiding judge on one case in their first year by constructing a panel with two senior or visiting judges. However, these deviations from strict randomness are small enough that federal judges themselves believe panels to be randomly constructed (Levy 2017).

We test the potential threat of nonrandom case assignment to identification across a number of dimensions by regressing our main independent variable, the fraction of a judge's co-panelists who are female in each year, onto a series of observed judge characteristics—specifically, on a judge's Hispanic ethnicity, quadratic of years of experience, quadratic of age, political party, quadratic of ideology of nominating president, and gender composition of current staff, controlling for judge gender and for court by year fixed effects. As shown in Table 1, there is little to no relationship between the exposure of a judge to female colleagues and any observed judge characteristics.

[TABLE 1]

Table 1 shows the relationship between a variety of judge characteristics and the main variable of interest, for both the full sample (columns 1 and 2) and separately for male and female judges (columns 3 and 4). We separate the sample by judge gender because the expected female share of colleagues is mechanically lower for female than for male judges, due to the fact that judges cannot interact with themselves. While we control for judge gender in column (2), the size of this mechanical effect is larger in small circuits such as the first circuit (with 10 judges) than in the ninth circuit (with 48 judges). Overall, the relationships we observe between judge characteristics and interaction with female colleagues is no greater than would be expected by chance, with the only statistically significant relationship being a lower likelihood of serving with female colleagues for female Hispanic judges at the 5% level. Because we perform 18 tests, a single test that is significant at the 5% level would be expected even if there were no true relationship between any of the judge characteristics and interactions with female colleagues.

5. Findings

5.1. Main Results

Table 2 presents ordinary least squares (OLS) regressions in which the dependent variable is an indicator of whether a judge hired a female clerk in year $t+1$ and the key independent variable is the fraction of the judge's co-panelists who were female in year t. Column (1) includes court by year fixed effects, with no additional covariates. Column (2) adds controls for judge gender, Hispanic ethnicity, and age, and column (3) adds controls for the political party of the judge's nominating president, a quadratic of the DW-Nominate score of the judge's nominating president, and a quadratic of the judge's years of experience on their current court. Table 2 shows that a one standard-deviation increase in a judge's exposure to female colleagues (an increase of 0.11 in the fraction of interactions with female colleagues), leads to a 4 percentage-point increase in the likelihood that a judge hires a female clerk. The addition of controls slightly increases the precision of the estimate but has negligible effect on its magnitude.

[TABLE 2]

We explore the robustness of the results by examining three potential sources of concern with our identification strategy. First, one might worry that judges' propensities to hire female clerks change over time in a manner that is correlated to their exposure to female judges. For example, if judges with greater seniority have more flexibility in scheduling vacations, experienced judges might schedule vacations while (predominantly less-experienced) female judges are working and hire their (predominantly male) favored clerk candidates. To test this hypothesis, we regress each judge's new staff in one year on that judge's exposure to female colleagues in the same year. Because staff are hired one to two years ahead of their start-date, this exposure cannot affect hiring. As shown in columns 1, 2, and 3 of Table 3, exposure to female clerks in the year following a hiring decision is unrelated to that hiring decision.[4]

Second, although unlikely given the structure of the appellate clerk market (see Section 2.1.1), one might worry that a judge's ideology and/or experience affects the labor pool from which they hire, either because judges prefer ideologically similar clerks or because more senior or more ideologically mainstream judges offer more prestigious positions and thus hire the most sought-

[4] Note that sample sizes for these placebo regressions are higher than for the primary regressions. Because hiring decisions happen two years prior to staffing starts, staff hired after 2014 cases are heard appear in our data in 2017, requiring us to drop case data from 2015-2017. In contrast, the placebo test only requires us to drop case data from 2007.

after clerks. Suppose the status of a particular group of judges (say, conservative judges) within a court covaries with the number of female clerks available within the appropriate local labor market (say, members of conservative judicial organizations), this could generate spurious correlation between the hiring of female clerks and the number of female co-panelists. We address this concern by matching each judge to the most similar judge within their court (with replacement) and regressing the exposure and characteristics of each judge to the hiring decisions of their match. We determine matches by regressing the fraction of each judge's staff who are female onto the judge's characteristics and court. We then select the judge with the most similar predicted staff gender composition and regress $Hire_{j,c,t+3}$ on $Inf_{k,c,t+3}$, where j is the reference judge and k is the match. As shown in columns 4, 5 and 6 of Table 3, while there is a positive association between the exposure of a judge's most similar colleague and their likelihood of hiring a woman, the relationship is substantially smaller than the actual estimated effects and not statistically significant.[5]

[TABLE 3]

Finally, we perform a permutation resampling procedure to determine whether our standard errors accurately reflect the distribution of likely effect sizes. To do this, we randomly reassign hiring decisions to judges within each court and year, and estimate the full model (shown in Table 2, column 3) for each random assignment. Figure 1 shows the distribution of 2000 randomly generated effects against the estimated effect, and shows that our results are unlikely to have occurred by chance.

[5] Note that differences in hiring pools across judges will only bias these regressions if they are correlated to differences within a court and year in exposure to female colleagues, which is unlikely. Were differences in hiring pools correlated to exposure to female colleagues, controls for judge characteristics would influence estimated results. As shown in Table 2, controlling for judge characteristics has a negligible impact on estimated effects.

[FIGURE 1]

5.2. Heterogeneity

We next examine heterogeneity in the effect of interaction with female colleagues. We examine seven sets of judge characteristics, examining whether judges with each characteristic are more affected by interactions with female colleagues and whether female colleagues with each characteristic have a greater effect when serving as co-panelists. These characteristics are: judge gender, judge quality, the fraction of a judge's current staff that are female, judge age, judge experience, the political party of the judge's nominating president, and whether the judge is visiting from a district court. We define judge "quality" as the rate at which a judge's decisions are cited, relative to other decisions from the same court and year. We find suggestive evidence that male judges, judges whose current clerks are less than 50% female, and judges with fewer average citations are more influenced by interaction with female colleagues than are female judges, judges whose current clerks are more than 50% female, and highly cited judges. These findings are consistent with a model of judge learning where judges who are most surprised by competent female colleagues have the largest changes in hiring.

[TABLE 4]

Panel A of Table 4 shows the relationship between a judge's characteristics and the effect of interaction with female colleagues on hiring. As shown in column 1 of Table 4, a 10 percentage-point increase in exposure to female colleagues increases the likelihood that a male judge hires at least one female clerk by 4.3 percentage points, but increases the likelihood that a female judge hires at least one female clerk by only 1.6 percentage points, though this difference may be the result of chance. Likewise, column 2 shows that judges whose current clerks are 50% female or more are less affected by interactions with female judges than are judges whose current clerks are less than 50% female. Together, these findings suggest that the judges most affected by interaction with female colleagues are those most likely to be surprised by interactions with capable female colleagues.

Panel B shows the relationship between the characteristics of female co-panelists and the effect of serving with the co-panelist on hiring. While none of these interactions are statistically significant, we find suggestive evidence that interactions with female district court judges have larger effects on hiring decisions than do interactions with female appellate judges. A 10 percentage-point increase in the fraction of co-panelists who are female district judges[6] increases the likelihood of hiring a female clerk by 22.2 percentage points, compared to a 3.4 percentage point effect of interacting with female appellate judges. The greater impact of interactions with female district judges may reflect the fact that district judges are less well known to appellate judges than are their appellate colleagues, so a given interaction with a district judge is likely to carry more information than would an interaction with an appellate judge.

6. Conclusion

This paper presents evidence that federal judges are more likely to hire female clerks after serving on a panel with female judges. We find that a one standard deviation increase in the fraction of published cases heard alongside female colleagues increases a judge's likelihood of hiring at least one female clerk by 3.5 to 4 percentage points. Because judges are broadly unconstrained in who they hire as a clerk, we interpret this change in hiring practices as a change in judge's assessment of the ability of women in judicial practice and law.

[6] As discussed in Section 2.1, district judges are not included in the analysis sample because their assignment to cases is nonrandom. However, a particular appellate judge's likelihood of being assigned to a case with a particular district judge is random.

This finding suggests that increases in the diversity of the upper rungs of a profession can shift attitudes in a way that creates opportunities at the entry level of the profession. This in turn suggests that policies aimed at increasing the diversity in the leadership of a profession, such as affirmative action policies or policies requiring that a certain number of board seats be filled by women, may have benefits beyond their immediate beneficiaries.

Bibliography

28 U.S. Code § 44. Appointment, tenure, residence and salary of circuit judges (https://www.law.cornell.edu/uscode/text/28/44)

Algan, Y., and Cahuc, P. 2006. "Job Protection: The Macho Hypothesis" Oxford Review of Economic Policy, 22(3), 390-410.

American Bar Association Required Disclosures. 2013. Enrollment and Degrees Awarded 1963 – 2012 Academic Years. (https://www.americanbar.org/content/dam/aba/administrative/legal_education_and_admissions_to_the_bar/statistics/enrollment_degrees_awarded.pdf)

American Bar Association. 2017. American Bar Association Standing Committee on the Federal Judiciary: What it is and How it Works. ISBN: 978-1-62722-880-0

Avery, Christopher and Christine Jolls and Richard A. Posner and Alvin E. Roth. 2007. "The New Market for Federal Judicial Law Clerks." University of Chicago Law Review 74, no. 2: 447–486.

Bagues, Manuel and Esteve-Volart, Berta. 2010. "Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment. " Review of Economic Studies, 2010, Vol. 77(4), 1301-1328

Bagues, Manuel and Labini, Mauro Sylos and Zinovyeva, Natalia. 2017. "Does the Gender Composition of Scientific Committees Matter?" American Economic Review, 2017, Vol. 107(4), 1207-1238

Beaman Lori and Raghabendra Chattopadhyay and Ester Duflo and Rohini Pande and Petia Topalova. 2009. "Powerful Women: Does Exposure Reduce Bias." Quarterly Journal of Economics, 124(4): 1497-1540.

Boisjoly, Johanne and Greg J. Duncan and Michael Kremer and Dan M. Levy and Jacque Eccles. 2006. "Empathy or Antipathy? The Impact of Diversity." American Economic Review, 96(5): 1890-1905.

Carrell, Scott and Lusher, Lester and Campbell, Duncan. 2018. "TAs Like Me: Racial Interactions between Graduate Teaching Assistants and Undergraduates," Journal of Public Economics, Volume 159, March 2018, Pages 203-224.

Chilton, Adam S. and Levy, Marin K.. 2015. "Challenging the Randomness of Panel Assignment in the Federal Courts of Appeals." Duke Law School Public Law and Legal Theory Series No. 2015-1; U of Chicago, Public Law Working Paper No. 529. Available at SSRN: https://ssrn.com/abstract=2520980 or http://dx.doi.org/10.2139/ssrn.2520980

Corno, Lucia and Eliana La Ferrara and Justine Burns. 2019. "Interaction, stereotypes and performance. Evidence from South Africa." IFS Working Papers W19/03, Institute for Fiscal Studies.

Dahl, G., Kotsadam, A. and Rooth, D.-O. 2018. "Does Integration Change Gender Attitudes? The Effect of Randomly Assigning Women to Traditionally Male Teams." Technical report, National Bureau of Economic Research, n.24351.

District Judges 28 U.S. Code § 292, 1986

Epstein, L. and Martin, A. D. and Segal, J. A. and Westerland, C. L. 2007. "The judicial common space." *Journal of Law, Economics, and Organization*, *23*(2), 303-325. https://doi.org/10.1093/jleo/ewm024

Federal Judicial Center. 2019. "Biographical Directory of Article III Federal Judges" https://www.fjc.gov/history/judges/biographical-directory-article-iii-federal-judges-export

Fortin, Nicole M. 2005. "Gender Role Attitudes and the Labour-market Outcomes of Women across OECD Countries. " *Oxford Review of Economic Policy*, Volume 21, Issue 3, Autumn , Pages 416–438, https://doi.org/10.1093/oxrep/gri024

Glynn, A. N. and Sen, M. 2015. "Identifying Judicial Empathy: Does Having Daughters Cause Judges to Rule for Women's Issues? " American Journal of Political Science, 59: 37–54. doi:10.1111/ajps.12118

Gramlich, J. 2018. "Federal judicial picks have become more contentious, and Trump's are no exception.", Pew Research Center.

Guiso, Luigi and Sapienza, Paola and Zingales, Luigi. 2003. " People's opium? Religion and economic attitudes. "Journal of Monetary Economics, 50, issue 1, p. 225-282, https://EconPapers.repec.org/RePEc:eee:moneco:v:50:y:2003:i:1:p:225-282.

History of the Federal Judiciary. 2019. http://www.fjc.gov. Web site of the Federal Judicial Center, Washington, DC.

Judicial Yellow Book: Who's Who in Federal and State Courts. New York, N.Y: Leadership Directories, Inc, 2007-2017. Publication Masters.

Leagle, 30 Nov. 2019, http://www.leagle.com/.

Levy, Marin K. 2017. "Panel Assignment in the Federal Courts of Appeals. " 103 *Cornell Law Review* 65-116

Levy, Marin K. 2019. "Visiting Judges." California Law Review, Vol. 107, No. 67, 2019; Duke Law School Public Law and Legal Theory Series No. 2019-11. Available at SSRN: https://ssrn.com/abstract=3328571

Lewis, Jeffrey B., Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin, and Luke Sonnet 2019. "Voteview: Congressional Roll-Call Votes Database." https://voteview.com/

Mark, Noah and R Harris, Daniel. 2012. " Roommate's race and the racial composition of white college students' ego networks." Social science research. 41. 331-42. 10.1016/j.ssresearch.2011.11.012.

McMillon, B. J. 2014. "US Circuit Court Judges: Profile of Professional Experiences Prior to Appointment." Technical report, Congressional Research Service.

NALP. 2019. " Clerkship Study Law Student Findings" https://www.nalp.org/clrkfind3

Pande, Rohini and Deanna Ford. 2011. " Gender Quotas and Female Leadership: A Review." mimeo, Harvard University.

Paluck, Elizabeth Levy and Green, Seth and Green, Donald P., 2017. " The Contact Hypothesis Re-evaluated." Avaliable at:
SSRN: https://ssrn.com/abstract=2973474 or http://dx.doi.org/10.2139/ssrn.2973474

Pérez, Israel Saeta, 2016. " Gender-guesser 0.4.0. " https://pypi.org/project/gender-guesser/

Posner, Richard and Roth, A. E. and Avery, Christopher and Jolls, Christine. 2001. "The Market for Federal Judicial Law Clerks." University of Chicago Law Review 68, no. 3: 793–902.

Reuben, Ernesto and Sapienza, Paola and Zingales, Luigi, 2014. "How stereotypes impair women's careers in science." Proceedings of the National Academy of Sciences Mar 2014, 111 (12) 4403-4408; DOI: 10.1073/pnas.1314788111

Rhinehart, Lynn K. 1994. "Is There Gender Bias in the Judicial Law Clerk Selection Process." 83 Geo L.J. 575

Stearns, Maxwell L. and Abramowicz, Michael. 2005. "Defining Dicta." *Faculty Scholarship*. 119.https://digitalcommons.law.umaryland.edu/fac_pubs/119

United States Courts. 2018. "Judicial Business, Table B-12." https://www.uscourts.gov/statistics/table/b-12/judicial-business/2018/09/30

Van Laar, Colette and Levin, Shana and Sinclair, Stacey and Sidanius, Jim. 2005. " The effect of university roommate contact on ethnic attitudes and behavior. " Journal of Experimental Social Psychology, Volume 41, Issue 4, Pages 329-345, ISSN 0022-1031, https://doi.org/10.1016/j.jesp.2004.08.002.

Washington, Ebonya L. 2008. "Female Socialization: How Daughters Affect Their Legislator Fathers." American Economic Review, 98(1): 311-32.

Tables and Figures

Table 1: Balance Tests for Random Assignment to Panels

Dependent variable: % co-panelists who are female	(1)	(2)	(3)	(4)
Age	0.0043 (0.0053)	0.0016 (0.0035)	0.0012 (0.0041)	0.0051 (0.0065)
Years on current court	0.0104** (0.0052)	-0.0008 (0.0032)	-0.0003 (0.0038)	-0.0004 (0.0064)
Ideology score	0.0343** (0.0169)	-0.0008 (0.0106)	-0.0043 (0.0126)	0.0188 (0.0254)
Republican	0.017 (0.0114)	0.0011 (0.0068)	-0.0014 (0.0079)	0.0118 (0.017)
% of current staff female	-0.0101 (0.0152)	0.0098 (0.011)	0.0185 (0.0137)	-0.0137 (0.0237)
Hispanic	0.0063 (0.0252)	-0.0148 (0.0124)	-0.0126 (0.0116)	-0.0748** (0.0341)
Court by year fixed effects	No	Yes	Yes	Yes
Sample	All	All	Male	Female
Observations	1074	1074	795	279

Notes: The table reports OLS estimation results from regressions of the fraction of co-panelists who were female in a year on a series of judge characteristics. Standard errors are robust and clustered at the judge level. Column (2) controls for whether the judge is female. Columns (2)-(4) include court by year fixed effects. Column (3) shows regression results for male judges, column (4) shows regression results for female judges. Significance levels are: * 10%, ** 5%, *** 1%. *Source:* Judicial yellow books, case dataset collected by authors (see data section for details).

Table 2: Effect of Serving with Female Judges on Hiring Decisions

Dep Var: Hired any female clerk in next year	(1)	(2)	(3)	(4)
Fraction of co-panelists who are female	0.3892**	0.4296**	0.3972**	0.4210**
	(0.1918)	(0.1867)	(0.1845)	(0.2036)
Female	0.0782**	0.0658	0.0352	0.0550
	(0.0395)	(0.0406)	(0.0384)	(0.0395)
Hispanic		0.2002***	0.1674***	0.1582***
		(0.0368)	(0.0464)	(0.0481)
Age		0.0141	0.0315	0.1156
		(0.1642)	(0.1817)	(0.2186)
Age^2		-0.0001	-0.0000	-0.0001
		(0.0001)	(0.0001)	(0.0002)
Years on current court			-0.0749	-0.0696
			(0.0617)	(0.0671)
Years on current court^2			0.0101	-0.0069
			(0.1227)	(0.1292)
Ideology score			-0.1774	-0.1385
			(0.1856)	(0.1943)
Ideology score^2			0.0437	0.0113
			(0.2453)	(0.2488)
Republican			0.0101	-0.0069
			(0.1227)	(0.1292)
% of current staff female				0.1721***
				(0.0527)
Other co-panelist characteristics	No	No	No	No
Court by year fixed effects	Yes	Yes	Yes	Yes
Observations	1074	1074	1074	987
Dependent variable mean	0.7030	0.7030	0.7030	0.6961

Notes: This table reports OLS estimation results from the regressions described in equations (1) and (2) in the text. Standard errors are robust and clustered at the judge level. The dependent variable is an indicator of whether a judge hired at least one female clerk in the following year, conditional on hiring any clerk. The table reports the results of regressions of the dependent variable on the fraction of co-panelists who were female in each year. Column (5) adds controls for the fraction of co-panelists who are Republican, who are younger than 60 years old, who have served fewer than 10 years on the court, who have a current staff that is more than 50% female, and who have an above average citation rate. Significance levels are: * 10%, ** 5%, *** 1%. *Source:* Judicial yellow books, case dataset collected by authors (see data section).

Table 3: Placebo Tests

	Prob. **Actual Judge** Hired any female clerks in **past** year			Prob. **Matched judge** hired any female clerks in **next year**		
	(1)	(2)	(3)	(4)	(5)	(6)
Fraction of co-panelists who are female	0.0070	0.1088	-0.0187	0.1777	0.1832	0.1353
	(0.1584)	(0.1614)	(0.1132)	(0.1810)	(0.1791)	(0.1818)
Female		0.0491	0.0205		0.0457	0.0343
		(0.0343)	(0.0281)		(0.0402)	(0.0377)
Hispanic		0.1663***	0.1086**		0.1823***	0.1538***
		(0.0357)	(0.0474)		(0.0368)	(0.0495)
Age		0.1093	-0.0170		0.0162	0.0982
		(0.1355)	(0.1161)		(0.1953)	(0.2167)
Age^2		-0.0001	0.0000		-0.0001	-0.0001
		(0.0001)	(0.0001)		(0.0002)	(0.0002)
Years on current court			-0.0530			-0.1499
			(0.0933)			(0.1916)
Ideology score			-0.0833**			-0.0726
			(0.0408)			(0.0671)
Ideology score^2			0.0186			0.0019
			(0.0574)			(0.1274)
Republican			-0.1679			0.0322
			(0.1710)			(0.2513)
Years on current court^2			0.0054			0.0065
			(0.0093)			(0.0181)
% of current staff female			0.9774***			0.1752***
			(0.0336)			(0.0524)
Court by year fixed effects	Yes	Yes	Yes	Yes	Yes	yes
Observations	1423	1423	1409	1046	1046	986
Dependent variable mean	0.70	0.70	0.70	0.70	0.70	0.70

Notes: This table reports OLS estimation results from placebo regressions. Standard errors are robust and clustered at the judge level. In columns 1-3, the dependent variable is an indicator of whether the judge hired at least one female clerk in year *t-1*. In columns 4-6, the dependent variable is an indicator of whether the most similar judge within a court, based on characteristics predicting the employment of female clerks, hired at least one female clerk in year *t+1*. The table reports the results of regressions of the dependent variable on the fraction of co-panelists who are female in each year. Judge characteristics include quadratics of judge age, experience in current position and ideology, judge gender, Hispanic ethnicity and party of nominating president. Significance levels are: * 10%, ** 5%, *** 1%. *Source:* Judicial yellow books, case dataset collected by authors (see data section).

Table 4: Heterogeneity in Main Effect

Dep Var: Hired any female clerk in next year

Panel A: Characteristics of judge

Var. Z:	Female (1)	>50% Fem Staff (2)	Abv-Avg Citations (3)	Republican (4)	Age < 60 (5)	< 10 yrs on court (6)	
Frac co-panelists female	0.4309** (0.2002)	0.4631** (0.2345)	0.4131** (0.1958)	0.2665 (0.2303)	0.3415* (0.2004)	0.2950 (0.2093)	
Frac co-panelists female X var. Z	-0.2720 (0.2858)	-0.2038 (0.2759)	-0.4161 (0.3445)	0.1402 (0.2842)	0.0186 (0.2534)	0.1747 (0.2599)	
Var. Z	0.0838 (0.0719)	0.1247* (0.0700)	0.1561* (0.0870)	-0.0478 (0.1514)	0.0677 (0.0807)	-0.0202 (0.0862)	
Court by year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	
Observations	1074	1074	1074	1074	1074	1074	
Dependent Variable Mean	0.7030	0.7030	0.7030	0.7030	0.7030	0.7030	

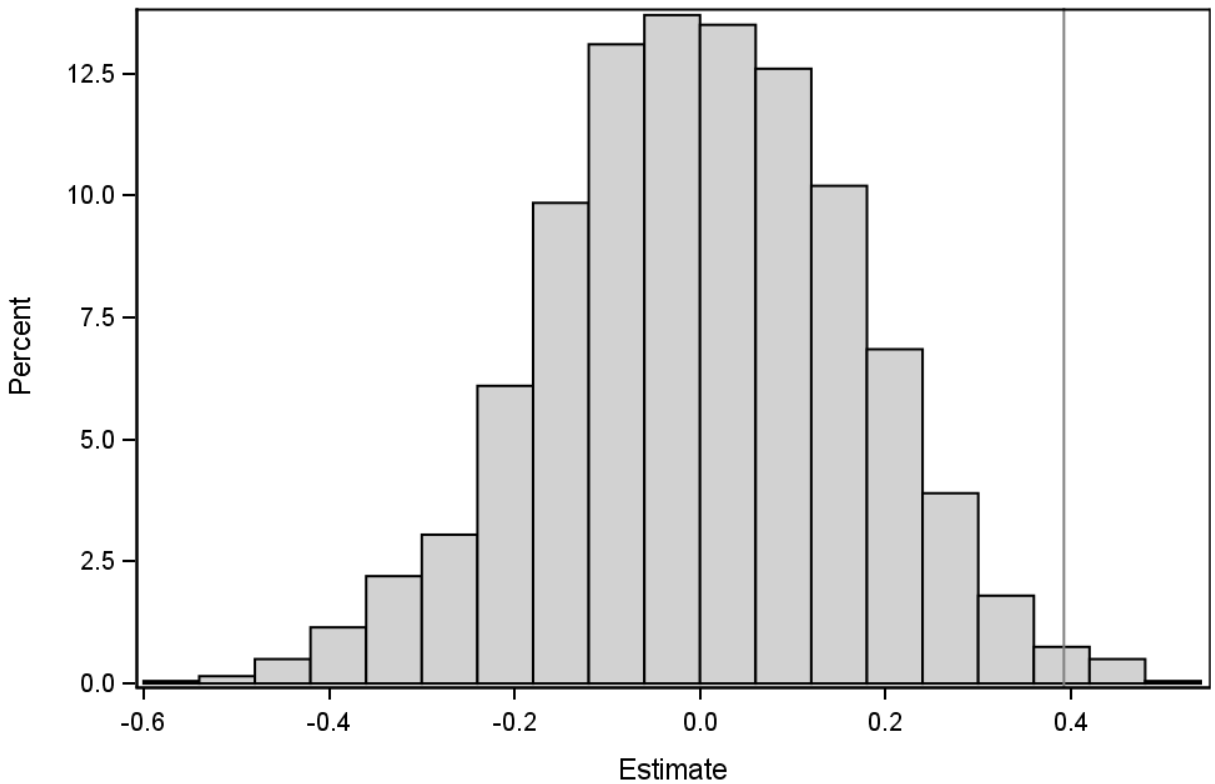
Panel B: Characteristics of co-panelists

Var. Z:		>50% Fem Staff (2)	Abv-Avg Citations (3)	Republican (4)	Age < 60 (5)	< 10 yrs on court (6)	District Judge (7)
Frac co-panelists female		0.4655** (0.2152)	0.3255 (0.2248)	0.5241*** (0.2013)	0.2872 (0.2854)	0.3788* (0.2177)	0.3376* (0.1839)
Frac co-panelists female AND var. Z = yes		-0.2280 (0.2441)	0.0799 (0.3476)	-0.4536 (0.3127)	0.0985 (0.3030)	-0.0790 (0.2956)	1.8873 (2.4469)
Court by year fixed effects		Yes	Yes	Yes	Yes	Yes	No
Observations		1074	987	1074	1074	1074	1074
Dependent Variable Mean		0.70298	0.69605	0.70298	0.70298	0.70298	0.70298

Notes: Panel A reports estimated coefficients on the interaction of the main independent variable with 6 interaction variables describing judge characteristics. Column (1) is an indicator of whether the judge is female. Column (2) is the fraction of the judge's staff that is female in year t, prior to the new hire. Column (3) is an indicator for whether the judge's decisions are cited more often than expected based on court and year. Column (4) is an indicator of whether the judge was appointed by a Republican president. Column (5) is the age of the judge (decades). Column (6) is the decades of experience of the judge. Column (7) is whether the judge is a district judge. Panel B reports the estimated coefficients on the interaction of the main independent variable with 6 characteristics of interacting judges. The main independent variable in each regression is the fraction of a judge's co-panelists who are female in each year. All results control for quadratics of judge age, experience in current position and ideology, judge gender, Hispanic ethnicity, party of nominating president, and the current fraction of the judge's staff who are female. Standard errors are robust and clustered at the judge level. Significance levels are: * 10%, ** 5%, *** 1%. Source: Judicial yellow books, case dataset collected by authors (see data section).

Figures

Figure 1: Randomization-Based Inference for Fraction of Co-Panelists who are Female



Notes: The figure shows distribution of coefficient estimates obtained from the final OLS specification in Table 2 while replacing the fraction of co-panelists who are female for a judge with the fraction of co-panelists who are female from a randomly selected judge from the same court and year. Vertical line represents the actual estimate obtained in the final specification in Table 2.

Online Appendix A: Detailed Data Description

A.1: Judicial Yellow Book

For each judge, the *Judicial Yellow Book* lists current courthouse, start date, date of birth, appointing president, and provides information on all staff members.[7] It particular, for each staff member the *Judicial Yellow Book* lists name, title, beginning and end of term, and contact information, and education. We use these data to determine which clerks worked for each judge in each year. We do so by taking advantage of the consistent formatting of the yellow book pages in the following way. Each subsection of the yellow books begins with a header "Chambers of Judge [judge's name]," followed by a list of judge characteristics and a list of staff, preceded by a "Staff" sub-header. Law clerks and other staff are then listed with their title, followed by their name, with biographical information indented on following lines. We use this formatting both to determine which staff work for each judge and to exclude staff with titles other than "Law Clerk," such as "Administrative Assistant" or "Judicial Assistant." We count 7,443 court clerks, with between 1 and 4 clerks working for each judge in each year in 98% of cases. The gender of judges and clerks are derived using the gender guesser algorithm (Pérez 2016), a tool that determines if a name is male, female or uncertain by comparing it to a database of 40,000 names from 54 countries, primarily in the United States and Europe. We have verified the results of this algorithm both by comparing its results to the relative frequency of men and women with each first name in the US Census and by confirming the genders of judges with ambiguous names through web searches. Using this method, we are able to determine the gender of 95.5% of clerks and 91% of judges. We

[7] The Yellow Books also contain information on the education (degrees earned, year of degree, and alma mater) and prior experience (in government, other judicial offices, law practice, private sector, military, and academia) of judges. In addition, they contain information on the education (degrees earned, year of degree, and alma mater) of judicial staff. This information, however, is largely incomplete, particularly for staff members, and is not used in our analysis. We use judge's college graduation year and law school graduation year to construct the age of the judge in case the date of birth is missing.

determined the gender of the remaining 9% of judges by examining biographical information from their court websites and/or news articles. Clerks for whom gender cannot be determined are excluded from the analysis (fewer than 5% of clerks).

Ethnicity of the judges is constructed by comparing judge names to surnames occurring at least 100 times in the 2010 decennial census. If more than 70% of census respondents with a judge's surname are Hispanic, we consider the judge to be Hispanic. If fewer than 30% of census respondents with a judge's surname are Hispanic, we consider the judge to be non-Hispanic. We determined the ethnicity of judges with an indeterminate surname through court websites and news articles.

A.2: Leagle.com Database

We determine which judges heard a case by exploiting the consistent formatting of court document headers to identify the area where presiding judges are typically listed. We then capture each word in this section of the document to determine whether it is a possible judge name or a linking or descriptive word, such as "Before," "Justice," or "Honorable". In following this procedure, we err on the side of including too many potential judge names rather than including too few. Therefore, we further compare the list of potential judge names to the list of surnames held by at least 100 people in the 2010 US Census, and remove all words not in the list of surnames.

In addition to these core variables, we also analyzed the text of the court's decision reported in the Leagle library to extract further information on the cases heard. This information is not used in the primary analysis, but contributes to our analysis of heterogeneity of effect. First, we extracted information on how many times each case was cited by the supreme court, appellate courts, and district courts. In addition to the verbatim decisions of each court, Leagle collects and attaches a list of cases that cite each included case. We use this information to count the citations of each case, and to categorize them by court and year. Next, we determine the decision writer for each case using consistent formatting of decisions within appellate circuits. We also collect information on whether a dissent was filed in each case, whether oral arguments were conducted, and whether an amicus brief was filed, but do not use these variables in our analysis[8]. We use these citation rates to construct a measure of "quality" for each judge, defined as the average number of Supreme Court and appellate court citations for cases published by the judge, as a proportion of the average rate of citation for cases published in the same circuit and year. While the citation rate of any particular case likely reflects the importance of that case as much or more as the quality of the decision writer, a high average citation rate is likely to indicate an ability to construct clear, convincing or novel legal arguments.

A.3: Other Data

In addition to our two primary datasets, we incorporate two additional sources of information: the ratings of judge quality collected by the American Bar Association (2017) and an indicator of judge ideology, as measured by the DW-Nominate of the judge's appointing president.

These ratings of judge quality are used in our analysis as a secondary measure of judge quality. We use this for two purposes: (1) to investigate whether a judge qualification affects the responsiveness of a judge's hiring decisions to exposure to female colleagues, and (2) to examine the effect of exposure to qualified female judges specifically (see Section 7). Each nominee to a

[8] We perform this textual analysis by taking advantage of the formatting of decisions and the presence of key words. We determine whether an amicus brief was cited in a case by searching the decision text for the term "amicus", and whether a concurring or dissenting was filed by searching for header text with the term "dissent, dissenting, dissents, concur, concurring, concurs" etc. We determine the decision writer using case formatting—opinions either begin with the decision writer's name or end with the writer's name. Citations are recorded in a standard bibliographic format, allowing us to both count citations and also determine which court and what level of court issued each citing opinion.

position in the federal judiciary is rated by the American Bar Association Standing Committee of the Federal Judiciary on the basis of their professional qualifications. According to the rules of the Standing Committee, ratings are made on the basis of a judge's "integrity, professional competence and judicial temperament," and do not reflect the judge's "philosophy, political affiliation or ideology" (American Bar Association 2017). This committee consists of 15 attorneys with standing to represent clients in appellate court circuits. Each member of the standing committee rates a nominee as either well-qualified,[9] qualified, or not qualified, and the committee reports both the opinion of the majority and the opinion of the next largest minority (American Bar Association 2017). We convert these ratings to numeric scores ranging from 0, meaning a unanimous rating of not qualified, to 5, a unanimous rating of well-qualified. Because 68 judges were confirmed prior to the online dissemination of American Bar Association ratings, we include 482 judges with qualification ratings.

We also include a measure of a judge's political ideology as both a control and as a source of heterogeneity in effect, taken from Epstein et. al. (2007). Presidents typically defer to senators from their own party on the nomination of a judge from a senator's state. Epstein et. al. (2007) exploit this senatorial courtesy to assign a judge the ideology of same-party home-state senators, when such exist, and the ideology of nominating presidents when home-state senators are of a different party than the president. The ideology of presidents and senators are based on the record of votes (for senators) and stated support (for presidents), calculated using the DW-Nominate algorithm (Lewis et. al. 2019).[10]

[9] In the 1989-1990 term, judges could also be rated Exceptionally Well-Qualified. We collapse this category with the Well-Qualified category.

[10] DW-Nominate scores are computed by examining the likelihood that each member of congress votes with each other member of congress.

Online Appendix Figures and Tables

Table A.1: Validation of Case and Judge Records Data

Court	Published cases in administrative records	Case records in database	% of records included in database	Number of judges: Judicial Yellow Books	Number of judges: Federal Judicial Center
	(1)	(2)	(3)	(4)	(5)
First Circuit	350	323	92.54	9.6	10.4
Second Circuit	310	301	95.25	22.4	24.3
Third Circuit	262	153	69.90	21.5	22.6
Fourth Circuit	205	192	94.17	13.9	17.4
Fifth Circuit	430	402	94.74	22.1	22.6
Sixth Circuit	364	350	95.94	28.1	29.1
Seventh Circuit	623	592	94.96	14.6	15.5
Eigth Circuit	608	575	94.11	20.4	20.1
Ninth Circuit	638	590	91.86	48.3	47.3
Tenth Circuit	309	274	89.11	19.8	21.4
Eleventh Circuit	280	271	97.38	16.0	17.1
District of Columbia Circuit	220	209	95.32	15.5	14.9
Federal Circuit	-	-	-	15.9	17.0
Total	4600	4232	91.99	268.0	279.6

Notes: Column (1) reports the number of published cases completed in each appellate circuit from 2007 to 2017, according to US Judicial Business Statistics. Column (2) reports the number of cases taken from the Leagle database from January 1, 2007 to December 31, 2017 collected by the authors. Column (3) gives the percent of cases reported in administrative data that are included in the Leagle database collected by the authors. Column (4) gives the average number of appellate judges included in the judicial yellow book data used in this paper in each year. Column (5) gives the average number of appellate judges reported by the Federal Judicial Center in each year.

Table A.2: Sample Selection

Sample	Observations	Appellate Judges
Appellate judges in Judicial Yellow Books, 2007-2017	2076	506
Appearing on at least one published appellate panel, 2007-2017	2164	298
Hired at least one clerk in year after appearing on panel	1074	215
With known gender	1074	215
With known race and age	1074	215
Appointed by President (non-magistrate judge)	1074	215
With known staff % female	987	197

Notes: This table reports the number of observations (judge X year) and the number of distinct appellate judges included in the analysis under each set of restrictions imposed on the data. The primary sample consists of 1074 observations from 215 appellate judges. *Source*: *Judicial Yellow Book*, case dataset collected by authors (see data section).

Table A.3: Characteristics of Judges and Clerks in Sample

	Full Sample Mean (SD) (1)	Male Mean (SD) (2)	Female Mean (SD) (3)	Full Sample Min / Max (4)	Sample Size Male / female (5)	M/F Diff (6)
Judge characteristics						
Female	0.2598 (0.4387)	0.0000 (0.0000)	1.0000 (0.0000)	0 / 1	795 / 279	***
Hispanic	0.0447 (0.2067)	0.0553 (0.2288)	0.0143 (0.1191)	0 / 1	795 / 279	***
Age (decades)	6.3389 (1.0125)	6.4479 (1.0304)	6.0283 (0.8910)	3.6 / 9.2	795 / 279	***
Decades on current court	1.5996 (1.0173)	1.7119 (1.0153)	1.2796 (0.9546)	-0.7 / 4.4	795 / 279	***
Ideology score	0.0635 (0.3612)	0.0869 (0.3540)	-0.0031 (0.3739)	-0.521 / 0.693	795 / 279	***
Republican	0.5279 (0.4995)	0.5660 (0.4959)	0.4194 (0.4943)	0 / 1	795 / 279	***
Number of clerks hired in year	2.8818 (1.2181)	2.8239 (1.2031)	3.0466 (1.2471)	1 / 7	795 / 279	***
Number of cases heard in year	56.2793 (39.4288)	56.4881 (39.9463)	55.6846 (37.9794)	1 / 194	795 / 279	
% of years with at least one female clerk hired[a]	0.7030 (0.4572)	0.6906 (0.4626)	0.7384 (0.4403)	0 / 1	795 / 279	
% of clerks that are female	0.4190 (0.4935)	0.4153 (0.4929)	0.4290 (0.4952)	0.0000 / 1.0000	2870 / 1035	

Notes: This table presents the average values of judge-level covariates in the analysis sample. For all variables other than % of clerks that are female, the sample is at the judge by year level. For % of clerks that are female, the sample is at the clerk by year level. *Source*: *Judicial Yellow Book*, case dataset collected by authors (see data section). a: of years where any clerk is hired

Table A.4: Raw and residual variation in female hires, female co-panelists

Panel A: fraction of co-panelists who are female	Mean	Std.dev.	Min	Max	Obs
Raw variable	0.23	0.11	0.00	1.00	1074
Residuals: net of court by year fixed effects	0.00	0.08	-0.24	0.77	1074
Residuals: net of court by year fixed effects, judge characteristics	0.00	0.08	-0.25	0.75	1074
Residuals: net of court by year fixed effects, judge fixed effects	0.00	0.06	-0.23	0.29	1074

Panel B: Hired a female clerk	Mean	Std.dev.	Min	Max	Obs
Raw variable	0.70	0.46	0.00	1.00	1074
Residuals: net of court by year fixed effects	0.00	0.43	-0.92	0.86	1074
Residuals: net of court by year fixed effects, judge characteristics	0.00	0.42	-0.92	0.80	1074
Residuals: net of court by year fixed effects, judge fixed effects	0.00	0.36	-0.96	0.86	1074

Notes: This table reports descriptive statistics for the key dependent and independent variables in this analysis. The key dependent variable is an indicator for whether a judge hired a female clerk in each year, conditional on hiring. The key independent variable is the fraction of a judge's co-panelists who were rated as highly qualified by a majority of American Bar Association raters who are female in each year. Judge characteristics include quadratics of judge age, experience in current position and ideology, judge gender, Hispanic ethnicity and party of nominating president. *Source*: *Judicial Yellow Book*, case dataset collected by authors (see data section).

Appendix Table A.5: Escalating interaction controls

Dep Var: Hired any female clerk in next year	(1)	(2)	(3)	(4)	(5)	(6)
Fraction of co-panelists who are female	0.3972** (0.1845)	0.3881** (0.1843)	0.3854** (0.1902)	0.3506* (0.1873)	0.4092** (0.1868)	0.3464* (0.1921)
Fraction of co-panelists who are Republican		-0.0571 (0.1422)				-0.1107 (0.1550)
Fraction of co-panelists <10 years in current position			0.2564 (0.1619)			0.1942 (0.2048)
Fraction of co-panelists <60 years old				0.2615* (0.1534)		0.1500 (0.1855)
Fraction of co-panelists above average citations					0.0562 (0.1438)	0.0121 (0.1491)
Judge Controls	No	Yes	Yes	Yes	Yes	Yes
Court by year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1074	1074	1074	1074	1074	1074
Dependent Variable Mean	0.7030	0.7030	0.7030	0.7030	0.7030	0.7030

Notes: This table reports OLS estimation results from the regressions described in equations (1) and (2) in the text. Standard errors are robust and clustered at the judge level. The dependent variable is an indicator of whether a judge hired at least one female clerk in the following year, conditional on hiring any clerk. The table reports the results of regression of the dependent variable on the fraction of co-panelists who were female in each year. Column (5) adds controls for the fraction of co-panelists who are Republican, who are younger than 60 years old, who have served fewer than 10 years on the court, who have a current staff that is more than 50% female, and who have an above average citation rate. Significance levels are: * 10%, ** 5%, *** 1%. *Source: Judicial Yellow Book*, case dataset collected by authors (see data section).