# Interactions with Powerful Female Colleagues Promote Diversity in Hiring

Marco Battaglini
Cornell University

Jorgen M. Harris
Occidental College

Eleonora Patacchini
Cornell University

**Abstract**

We study 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 judicial panels at random and have few limitations on their choices of law clerks. Using a unique dataset of federal case records merged with judicial hiring information, we find a significant effect of the fraction of co-panelists who are female on a male judge's likelihood of hiring a female clerk. This finding suggests that increases in the diversity of the upper rungs of a profession can create opportunities at the entry-level.

This work was supported by a grant from the Cornell Institute for Social Sciences (ISS_ATF_F17). We would also like to thank our fantastic team of undergraduate and post-graduate research assistants. Francesca Guiso, Sam Markiewitz, Becky Cardinali, Cenjia Lu, Ignasi Merediz Sola, Linchen Zhang, and Hongyi Duan: thank you for your hard work, creativity, and patience.

## 1. Introduction

Workplaces are segregated by gender, with women working in lower-paid occupations (Blau & Kahn, 2017) and firms (Goldin, Kerr, Olivetti, & Barth, 2017) than do men with similar levels of education. As a consequence, men are less likely to work closely with women than with other men, particularly in high-paying fields. A longstanding literature argues that this segregation and ensuing lack of exposure contributes to discrimination against women in male-dominated professions by reinforcing negative perceptions of women's ability (Allport, 1954), reducing the quality of information that men receive about female colleagues (Aigner & Cain, 1977), and socially reinforcing taste-based discrimination. In this case, promoting diversity in professional interactions may improve economic opportunities for women, racial minorities, and other "out-groups" by changing the attitudes and behavior of men and other "in-group" decision-makers. Empirical validation of this hypothesis has important policy implications, as it suggests and legitimizes practical steps to promote diversity in the workplace.

Tests of the effect of exposure to "out-groups" on the attitudes and behavior of "in-groups" face a key challenge: individuals with negative beliefs about out-groups are less likely to meaningfully interact with members of out-groups (Bertrand & Duflo, 2016). As a result, credible research designs have focused either on experimental evidence or on the rare circumstances when people interact with others at random. For instance, seemingly random assignment to college roommates (Boisjoly et. al., 2006) or to gender-integrated military squadrons (Dahl, Kotsadam, & Rooth, 2021) influences attitudes toward women and ethnic minorities.

While this literature suggests that contact with out-groups can change attitudes and behavior, prior work has either examined inexperienced people at the beginnings of their careers, or examined intimate family relationships. No previous work has credibly examined the effect of

ordinary workplace interactions on high-stakes decisions, particularly on mature, well-established professionals in the elite of their occupations.

In this paper, we exploit a unique feature of the Federal appellate court system to present clear causal evidence on the effect of workplace interactions with out-groups on individual hiring decisions. Appellate court judges hear cases in panels of three randomly selected judges. Because judges hear cases at random, their likelihood of working with female colleagues on cases is also random. At the same time, appellate judges are broadly unconstrained in who they hire as a court clerk, a prestigious position typically filled by graduates of top law programs. We thus examine the effect of exogenously determined interactions with women on appellate panels on a judge's choice of court clerk.

We find significant positive effects on the likelihood of hiring female clerks from professional interactions with female judges. In particular, we find that a one standard deviation increase in the fraction of case interactions a male judge carries out with female colleagues increases the likelihood of hiring at least one female clerk in the next year by 7.8 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.

Our work relates to three strands of literature. The first literature examines the effect of exposure to out-groups on attitudes and decisions. Experimental psychologists have extensively studied these effects, starting with Allport's (1954) development of the contact hypothesis, and economists have more recently addressed this question experimentally and quasi-experimentally.[1] As previously mentioned, prior work has focused on college students and other new entrants to

[1] For a review of recent work, see Paluck, Green, and Green (2018).

professions. Boisjoly et al. (2006) use the random allocation of college roommates to show that inter-racial interactions make the advantaged group more empathetic to the disadvantaged group while improving the academic performance of the disadvantaged group. Similarly, Dahl, Kotsadam & Rooth (2021) find that random assignment to gender-integrated boot camp squads leads men to adopt more egalitarian attitudes and to enter military occupations with more female representation.

Work focused on older adults has primarily considered interactions outside of the workplace. Mousa (2020) and Lowe (2021) examine the effect of random assignment to soccer and cricket teams with Muslim or other-caste teammates. Both studies find that players assigned to out-group teammates are more willing to join mixed teams in future seasons, more likely to vote for out-group players for an award, and more likely to report out-group friendships. Other work examines the effect of close family relationships on decisions. Washington (2008) finds that lawmakers with daughters vote in a more liberal manner, especially on reproductive issues. While these papers provide strong evidence that exposure to out-groups shapes choices, close personal relationships may have very different effects on behavior than do working relationships. Furthermore, none of these papers examine decisions related to hiring or promotion.

We contribute to this literature by proposing quasi-experimental evidence on how professional interactions with female colleagues affect an important economic decision: the hiring of court clerks, a prestigious position filled by top law students. Our approach is unique because it directly estimates the effect of professional interactions on hiring decisions. Examining the upper reaches of the job market is especially important since this is where the gender wage gap is the largest (Blau & Kahn, 2017).

We contribute to a second literature examining the effect of policies that increase the number of women in leadership positions on opportunities for women. In contrast to work on interaction with out-groups, this work has mostly found small or no effects on economic opportunities for women. Bertrand et. al. (2019) examine a Norwegian policy mandating that at least 40% of the board members of public companies are women, finding no effect on the number of women hired to the top echelons of public Norwegian companies, nor on the gender earnings gap. Likewise, Maida & Weber (2020) find that a similar reform in Italy increased the number of women on boards but had small and imprecisely estimated effects on the number of women hired as senior managers. Beaman et. al. (2009) find somewhat more positive effects of an electoral law in West Bengal requiring that a third of village council seats be held by women. In precincts whose seats are reserved for women, male voters are more likely to elect women for open seats despite expressing stronger preferences for male leaders.

Our paper is similar to this literature in that it examines the effect of exogenous workplace interaction with elite female peers on opportunities for less elite women. However, the context of the legal profession differs from that of corporate boards in a few key ways. As noted by Bertrand et. al. (2019), boards of directors have limited influence over hiring decisions. As a result, female representation on boards may not affect hiring even if it did affect the beliefs, attitudes, or preferences of male board members. In contrast, appellate judges have broad and unconstrained authority to hire law clerks. Furthermore, boards make decisions as a collective deliberative body. As a result, changes to board decisions reflect both changes to male board members' preferences and the preferences of female board members. As demonstrated by Bauges & Esteve-Volart (2010), majority-female committees may overestimate the qualifications of male job candidates in

some circumstances, potentially counterbalancing increased willingness to hire women among male board members. In contrast, appellate judges make hiring decisions unilaterally.

Serving with female board members may also affect men's attitudes and beliefs differently than does hearing cases with female judges. Because female board members are chosen by existing board members, they may already be well-known to their male colleagues, limiting the amount that men learn by interacting with them. For instance, Chevrot-Bianco (2021) finds that a third of women appointed to Danish boards of directors following the imposition of a gender quota are relatives of male board members. Because female appellate judges are not chosen by male colleagues, men may learn more by interacting with women in the judiciary. In addition, legally mandated gender quotas may inspire backlash against female professionals among male board members, counteracting any positive effects of interaction with women. While interaction with women in the judiciary is nonvoluntary, random and gender-neutral assignment to peers may not inspire backlash.

Finally, our work contributes to a literature studying the decisions of judges. Boyd, Epstein & Martin (2010) examine the effect of hearing cases alongside female colleagues on a judge's rulings, but have not examined decisions made outside of court, such as hiring. Glynn & Sen (2015) examine the gender of judges' children on their voting decisions. Ash, Chen, & Ornaghi (2020) examine the role of gender bias, measured by the use of sexist language in written opinions, on judge's rulings, likelihood of reversing female district judges, and citation of female authors. Our study is relevant in this context because provides insight on how the judges' professional interactions affect their decisions outside of court.

## 2. Institutional Setting

### 2.1. Federal Appellate Courts

In the U.S. Federal court system, cases originate in district courts, where parties present evidence and individual judges make an initial ruling. Each party is entitled to appeal decisions to appellate courts, which review the legal reasoning used in district courts. Since the Supreme Court is not required to review an appellate court's decision, the great majority of appellate court decisions are final.

Twelve federal appellate courts hear cases decided by district courts in a distinct geographical area, while the thirteenth hears cases involving patents, international trade, and other particular subject areas originating anywhere in the United States. The geographical organization of the courts requires federal judges to hear cases on all subjects, requiring them to maintain expertise in a wide variety of legal areas. Therefore, a judge might learn different information about a colleague's ability in each case that they hear with them.

Appellate judges hear cases in panels of three, but occasionally organize larger panels. Each court assigns judges to cases using methods intended to be random, such as the use of computer programs and the drawing of lots (Levy, 2017).[2]

Courts construct panels with regular appellate judges, senior appellate judges (who work part time), and visiting judges—typically either district judges or retired federal judges. Most courts ensure that there are at least two regular appellate judges on each case, so visiting judges hear cases with other visiting judges only in exceptional circumstances (Levy, 2019). In addition,

[2] The algorithms used by some courts, such as the Fifth Circuit, avoid having any judge serve too often with any other judge in a year (Levy, 2017). This limits the variance in our sample but preserves random assignment to colleagues.

chief justices often restrict the type of cases that visiting judges hear, for instance only using them for civil cases (Levy, 2019). Furthermore, because visiting judges only hear cases for part of the year, the timing of their visits affects the colleagues they interact with. As a result, while neither regular appellate nor visiting judges are assigned to cases on the basis of their characteristics, visiting judges hear a different set of cases with a different set of colleagues than do regular appellate judges. For this reason, we only consider the hiring decisions of appellate and senior appellate judges. However, when measuring exposure to female colleagues, we consider the entire population of judges, including visiting judges, as explained in Section 3.1.

Judicial panels decide whether to affirm or overrule the decision of the appellate court by considering legal briefs prepared by each party. A majority (typically two members) of a court must agree to the court's decision, and in 87% of cases, the court reaches a unanimous decision. If the judges believe that the case addresses novel legal reasoning and will thus be useful as precedent, they publish the opinions, making them available in legal registers and online databases.

### 2.1.1. Appointment of Judges and Clerks

Judicial appointees are nominated to lifetime appointments by the President and confirmed by the Senate.[3] Most judges in our sample are over 60 years old, and have served in their current positions for more than 15 years. As a result, they entered a legal profession with far fewer women than are present today. While women make up half of new lawyers today, fewer than 5% of new lawyers were women prior to 1968, 36% were women in 1981, the year that the median federal judge in our sample finished law school, and 43% were women when the median judge started their current position (American Bar Association, 2017). As a result, federal judges hire in a labor

[3] An exception is magistrate judges—district judges appointed by current members of a district court (McCabe, 2014).

market where women are significantly more numerous and successful than they were when the judges first formed impressions of the legal profession and of the judiciary.

Each judge hires a staff of administrative assistants and law clerks. Typically hired directly out of law school, law clerks typically serve one to two-year terms assisting judges with legal research and decision writing (Posner et. al, 2001). Appellate court clerkships are prestigious and highly sought-after positions, and lead to positions at top law firms, government agencies and the judiciary (Rhinehart, 1994). They are especially valuable for law students hoping to enter the judiciary—every Supreme Court clerk serving from 2005-2014 had previously been an appellate court clerk (Hess, 2015).

Most law clerks accept clerkships in their second year of law school (Avery et. al, 2007)—two years before their start-date. Given this, we define a clerk's year of hiring as two years prior to the year that they begin work.[4]

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 law students apply to clerkships, judges receive thousands of applications (N.A.L.P., 2019). Judges typically give very short decision windows to accepted applicants (in some cases, as brief as 10 minutes), so most law students take the first clerkship offered.

## 3. Data Description

We pool data from several sources. Our two primary datasets are the *Judicial Yellow Book*—a directory published annually by Leadership Directories Inc. that lists information on judges and clerks working in the US court system (Leadership Directories, inc., 2007-2017), and a directory of case records published by federal appellate courts from 2007 to 2017 (Leagle, 2019).

[4] Thus, clerks hired in year $t+1$ begin work three years after year $t$.

### 3.1. Primary Data

We collect judge and clerk information from the *Judicial Yellow Book*, which includes the names and backgrounds of judges serving at all levels of the federal court system, and 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 as described in detail in Appendix A.

To determine which judges sat on panels together in each year, we scraped information from the online court records aggregator Leagle.com. Leagle stores and categorizes the decisions handed down by the United States courts – trial courts, appellate courts, and the Supreme Court. The library is comprehensive and contains over 5 million cases since 1950. We pool information on the universe of published cases heard between 2004 and 2017, in total 50,813 cases. For each case, Leagle provides the full text of the court's decision, exactly as it appears in published court documents. We use this text to determine which judges heard each case and to collect information on the complexity of each case, as described in Appendix A.

We combine these two datasets by matching judge names appearing on each appellate case to judge names listed in the *Judicial Yellow Book* in the appellate court that heard the case or a subsidiary district court.[5] When merging the case data to the list of judges in the *Judicial Yellow Book*, 66% of potential judge names identified in the case records match the relevant appellate court or subsidiary district courts. The remaining 34% consist of attorneys, parties incorrectly categorized as judges, and in fewer cases, judges visiting from other circuits and retired judges

[5] We include judges from subsidiary district courts because they occasionally serve on appellate panels as visiting judges (28 U.S. Code § 292).

hearing cases as senior judges, and are thus dropped from the analysis. Among appellate judges listed in the *Judicial Yellow Books*, 85% appear on at least one published case in the year that they are listed. The remaining 15% are mostly senior judges who hear few or no cases each year. Among district judges listed in the *Judicial Yellow Books*, 12% appear on at least one published case in the year that they are listed, consistent with a significant minority of district judges visiting appellate courts each year.

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 2004-2017. In our final database of 50,484 cases, 70% have three judges that are included in the analysis, 20% have two judges, 6% have one judge, and 4% have more than three judges. Cases can have fewer than three judges when they include visiting judge who is not from a subsidiary district court or when names are recorded improperly in our database. 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.

Appendix Table A.1 compares the data on judges and cases used in this paper to official court statistics. First, we compare the number of published cases in our data with the number reported in the annual Judicial Business Statistics (Judicial Business: Table B-12). Because our data source covers January 2007-December 2017 whereas the Judicial Business Tables covers November 2006 to November 2017, we expect small deviations in the number of cases across the two datasets. Appendix Table A.1 shows that the number of cases in our data are remarkably similar to those reported in official data, with the exception of the Third Circuit, where we recover fewer cases than expected. 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 Judiciary, 2019) serving in each year. An average of 267 appellate judges appear in each year of our data, compared to 280 appellate judges in each year of the federal judicial center data. This discrepancy is due primarily to the exclusion of some senior appellate judges from *The Judicial Yellow Book.* In particular the Judicial Center lists several senior appellate judges in the Fourth circuit 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 also do not appear in our case records data, we believe that these judges have maintained senior status but are not actively hearing cases.

In addition to these core variables, we also analyzed the text of the court's decision reported in the Leagle library to determine the number of cases cited by each opinion, the length of each opinion, whether oral arguments were conducted, and whether an amicus brief was filed. We use this additional information to demonstrate that published cases are of a higher complexity than unpublished cases.

### 3.2 Other Data

In addition to our two primary datasets, we incorporate two additional sources of information: The Biographical Directory of Article III Judges, compiled by the Federal Judicial Center, and a rating of the conservatism of the president who appointed each judge using the DW Nominate algorithm (Epstein, Martin, Segal, & Westerland, 2007). We provide further information in Appendix A.

### 3.3. Sample Selection

We construct a panel dataset where an observation represents an appellate court judge in a particular year. As shown in Appendix Table A.2, our sample includes years when a judge heard at least one published case and hired at least one clerk in the following year. This sample includes 1339 observations from 229 judges over ten years. Our primary analysis examines the 996 observations from 174 male judges. Some specifications control for the current gender composition of a judge's law clerks—this covariate is missing for 77 observations where a judge has no law clerks on staff, primarily in the first year of a judge's tenure. We provide further details in Appendix A.4.

Table 1 shows characteristics of this sample. As shown in Table 1, 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 69% of years in which they make a hiring decision. Overall, 42% of appellate court clerks are women, substantially lower than the 59% of district court clerks who are women.

## 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 at least one female clerk in the following year. Because assignment of appellate judges to panels is random conditional on court and year, we control for fixed effects at the court by year level. We estimate the effect of exposure to female colleagues on the likelihood of hiring a female clerk using the following regression equation:

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

where $Hire_{j,c,t+1}$ is an indicator of whether judge $j$, in court $c$, hired at least one female clerk in year $t+1$; $Inf_{j,c,t}$ is exposure to female judges in year $t$, $X_{j,t}$ is a set of judge characteristics, and $\theta_{c,t}$

is a set of court by year fixed effects. We examine the hiring of at least one female clerk because we expect the effect of exposure to female colleagues to be greatest for judges who would not otherwise hire a female clerk. As a robustness check, we also examine effects on the fraction of hired clerks who are female and the number of hired clerks who are female in Section 5.2.

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. Because we use the fraction of co-panelists who are female, rather than the count, our measure does not depend on the volume of cases heard by a judge in a particular year. This decision reflects the expectation that full-time judges with few cases have more time-consuming cases than those with many cases, and that each case is more salient when judges hear fewer cases. We primarily examine the hiring decisions of male colleagues because we expect men to be more likely to change beliefs and preferences about women in response to interaction with female colleagues.[6]

To demonstrate the feasibility of this approach, we show that sufficient variation is left after accounting for our large set of controls. As shown in Appendix Table A.3, more than half of the variation in the hiring decisions of judges and the exposure of judges to female colleagues remains after accounting for judge characteristics, court by year fixed-effects, and judge fixed-effects.

### 4.1 Evidence in Support of Identification Strategy

This paper's key identifying assumption 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

[6] As shown in Appendix Table A.7, we find statistically insignificant effects on hiring decisions of female judges, with a point estimate that is just over half the effect on male judges We also examine effects on all judges as a robustness check, finding smaller but highly statistically significant effects.

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 M. & Abramowicz, 2005). While violations of pure randomness are inevitable, these violations 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, work by Chilton & Levy (2015) finds 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. Levy (2017) examines a broader range of potentially non-random scheduling decisions made by the chief justice's office of each appellate circuit, finding, 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 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: the judge's racial and ethnic background, years of experience, age, political party, the ideology of their nominating president, and the gender composition of their current staff, controlling for judge gender and for court by year fixed effects.

Table 2 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). Overall, the observed relationships between judge characteristics and interaction with female colleagues are no greater than would be expected by chance. One correlation is statistically significant at the 10% level: men's interaction with female colleagues is negatively associated with the fraction of their staff that is female. Because we perform 24 tests, a single test that is significant at the 10% level would be expected even if there were no true relationship between any of the judge characteristics and interactions with female colleagues. We confirm this with a test of joint significance for each group of eight regressions, none of which rejects the hypothesis that all estimated effects are zero. [7]

We conclude this section by documenting the differences between published and unpublished cases. As discussed in Section 2.1, we use data from published cases. We do so both because only published cases are systematically available in all courts and years and because published cases involve more and higher-stakes interaction with co-panelists than do routine, non-precedential cases. As a result, published cases provide more information to co-panelists about a judge's abilities.

Because the decision to publish a case is made by the judges empaneled to hear the case, it is important to assure that the assignment of a female judge to a case does not affect its likelihood of being published. In particular, one might conjecture that judges who consider women to be less capable than men will be less likely to support publication when seated on a panel with female judges. Were this the case, we would underestimate the exposure of sexist judges to female colleagues, positively biasing our regressions.

[7] As we show in Section 5.1, controls for judge characteristics, including party, do not weaken estimated results. As shown in Appendix Figure A.1, results are robust to the exclusion of each court, each judge, and year.

We explore this hypothesis by examining cases from 2016 and 2017, the only years in our sample in which all appellate circuits made a substantial number of unpublished cases available online. We first examine whether women publish a smaller fraction of cases than do men, as would be expected if some judges were reluctant to publish when empaneled with women. We regress the fraction of cases on which a judge is empaneled that are published in each year on a judge's gender, under the regression framework presented in equation 1. As shown in Appendix Table A.4, we find that cases heard by female judges are published at a 2% higher rate than are cases heard by male judges, a difference that is neither statistically significant nor meaningful as potential sources of bias. This indicates that on average, judges are no more or less likely to agree to publish a case when empaneled with a female colleague.

Next, we examine whether judges with potential sexist beliefs are less likely to publish cases when empaneled with female colleagues. We calculate the ratio of female colleagues on published case to female colleagues on unpublished cases. Were sexist judges less likely to publish a case when empaneled with female judges, they would have fewer female co-panelists on published cases and more female co-panelists on unpublished cases than would other judges. As a result, we would expect this ratio to be positively correlated with judge's preferences for female clerks if judges published selectively, but uncorrelated with judge's preferences for female clerks otherwise. We regress this ratio onto the fraction of clerks hired by a judge in all observed years who are female ($AFF_{j,c,t}$, defined in Section 3.4). As shown in Appendix Table A.5, the relationship between publication ratio and average female fraction of clerks is small, statistically insignificant, and in the opposite direction predicted by an effect of sexist beliefs on publication decisions.

In contrast, we find clear evidence that published cases differ substantively from unpublished cases along other observed characteristics. As reported in Appendix Table A.6, Amicus briefs—legal arguments filed by people or organizations not party to a case—were filed for 9.4% of published cases, compared to 0.3% of unpublished cases. This suggests that published cases are far more likely than unpublished cases to set precedents relevant to outside parties. Likewise, judges requested oral arguments for 37% of published cases, compared to 6% of unpublished cases. On average, published opinions were nearly four times the length of unpublished opinions and included more than three times as many citations. In sum, our evidence indicates the publication probability depends on factors unrelated to the gender of the panel of judges: men and women are equally likely to be on published cases.

## 5. Findings

### 5.1. Main Results

Table 3 shows the ordinary least squares estimates of equation 1, 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, a quadratic of age, 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. Column 3 adds judge fixed effects. Column 4 controls for the number of clerks hired by the judge in year $t+1$ while column 5 controls for the fraction of a judge's staff that is female in year $t$. Because the addition of this control excludes judges in years where they have no existing staff, we consider column 4 to be our main

specification. Consistent with our identification strategy, differences in these specifications are no larger than would be expected by chance.[8] Column 4 shows that a one standard-deviation increase in a male judge's exposure to female colleagues, or an increase of 0.11 in the fraction of judicial interactions with female colleagues, leads to a 7.8 percentage-point increase in the likelihood that a judge hires a female clerk.

We also examine the effect of exposure to female colleagues on the number of women hired in year *t+1* and the fraction of new hires who are women in year *t+1*. Our results are qualitatively similar using these alternative dependent variables. As shown in Appendix Table A.8, the effect of exposure to female colleagues on the number of women indicates that a one standard deviation increase in exposure to female colleagues increasing the number of women hired by 1.03, while the effect on the fraction of hires who are women is 5.0 percentage points.[9] Results using all three dependent variables indicate a sizeable effect of exposure to female colleagues on the hiring of women.

### 5.2. Robustness

We conduct two exercises to further increase our confidence that judges' exposure to female clerks is unrelated to unobserved judge characteristics.[10] First, we regress an indicator of

[8] We conduct F-tests for equality of coefficients for our primary independent variable comparing each column to each other column. We cannot reject the hypothesis that differences in the estimated coefficients occurred by chance in any of these comparisons. The largest difference is between columns 1 and 3, with a p-value of 0.5844.
[9] There is no effect of exposure to female colleagues on the total number of clerks hired. The increase in the number of women hired corresponds to a commensurate decrease in the number of men hired. The effect on the fraction of clerks who are women is smaller than the effect on hiring at least one woman (7.8 percentage points, shown in column 4 of Table 3) because the effect on hiring more than one woman is small and imprecisely estimated but not zero.
[10] Exposure to female judges could only be related to judge characteristics as a consequence of small deviations from perfect randomization in the assignment of judges to cases. For example, if judges with greater influence have more flexibility in scheduling vacation days, a judge might take vacations that are scheduled while (predominantly non-senior) female judges are working. Greater seniority might also increase their likelihood of hiring their (predominantly male) favored clerk candidates.

whether judge *j* hired any female clerks in year *t-1* on judge *j*'s exposure to female colleagues in year *t*. Because staff are hired two years in advance of their start-date, exposure to female colleagues in year *t* cannot have a causal effect on hiring in year *t-1*. As shown in columns 1, 2, and 3 of Table 4, exposure to female clerks in the year following a hiring decision is unrelated to that hiring decision, providing evidence in support of our identification strategy.[11]

Second, we match each judge *j* to the most similar judge within their court (with replacement). 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 $Hire_{j,c,t+1}$ on $Inf_{k,c,t+1}$, where *j* is the reference judge and *k* is the match. As shown in columns 4, 5 and 6 of Table 4, the association between the exposure of a judge's most similar colleague and the judge's likelihood of hiring a woman is negative and not statistically significant.[12]

Apart from questions over the random assignment of judges to co-panelists, one might worry that judges hire more female clerks when interacting with female colleagues due to characteristics of their colleagues other than gender. For instance, if judges are more likely to hire liberal clerks after interacting with liberal colleagues, exposure to female colleagues would be associated with the hiring of female clerks. We address this concern by including additional controls to our main regression for the fraction of co-panelists who are Republican, who have

[11] Note that the sample sizes for these placebo regressions are higher than the sample sizes 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.

[12] This exercise is a further check against the possibility that our results are driven by unobserved shocks affecting particular subsets of judges serving on an appellate circuit. Factors correlated with a judge's ideology and/or experience might affect the labor pool from which they hire, either because judges prefer ideologically similar clerks or because more senior or more ideologically mainstream judges can offer more prestigious positions and are thus able to hire the most sought-after clerks. If a quirk of the assignment of judges to cases led similar judges to be more likely to hear cases with female colleagues in some years, this could produce spurious correlations between exposure to female colleagues and hiring.

served on the court for fewer than 10 years, who are younger than 60, and who have above-average rates of citation to their published opinions. As shown in Appendix Table A.9, none of these other judge characteristics significantly mediate the relationship between the gender of co-panelists and the gender of newly-hired clerks.

Additionally, as discussed in Section 3.1, case data from the Third Circuit is likely to be incomplete. Likewise, as discussed in Section 4.1, small deviations from true random assignment have been observed in the Second, Sixth, Ninth, and DC Circuits. In order to ensure that our results are not driven by a single circuit, we re-estimate our regressions, iteratively omitting data from each circuit. Similarly, we iteratively omit data from each hiring year and from each judge. As shown in Appendix Table A.10, our results are not driven by any single circuit or year. Our estimated effect is smallest when omitting the District of Columbia circuit, at 0.59 and when excluding the 2009 hiring year, at 0.54. Likewise, as shown in Appendix Figure A.1, our results are not driven by a single judge, with estimates excluding a judge ranging from 0.58 to 0.69. Estimates are statistically significant at the 0.01 level in all cases.

Finally, we perform a permutation resampling procedure to demonstrate that the results shown in Table 3 are highly unlikely to have occurred by chance. To do this, we randomly reassign exposure to female colleagues to judges within each court and year, and estimate the full model (shown in Table 3, column 4) for each random assignment. Figure 1 shows the distribution of 10,000 randomly generated effects against the estimated effect. The figure shows that while those estimates are small and centered around zero, our estimate, shown by the gray, vertical line, is in the upper tail of the distribution, leaving on the right only one random estimate, or 0.01% of the cases.

### 5.3. Additional Evidence

We estimate event study regressions for two purposes. First, we expand on the falsification test of our empirical strategy in Table 4 by examining the effect of exposure to female colleagues in year *t* on hiring decisions made *during and prior to year t*. Next, we explore whether interactions with female colleagues has a persistent effect on judge's hiring decisions by examining the effect of exposure to female colleagues in year *t* on hiring decisions made in years *subsequent to year t+1*. Specifically, we estimate the following regression equations, with values of *k* ranging from -5 to +3, where $Hire_{j,c,t+k}$ is an indicator of whether judge *j*, in court *c*, hired at least one female clerk in year $t + k$; $Inf_{j,c,t}$ is exposure to female judges in year *t*, $X_{j,t}$ is a set of judge characteristics, and $\theta_{c,t}$ is a set of court by year fixed effects.[13]

$$Hire_{j,c,t+\mathrm{k}} = \beta_k Inf_{j,c,t} + \delta X_{j,c,t} + \theta_{c,t} + \varepsilon_{j,c,t+k} \qquad (2)$$

The OLS-estimated coefficients estimated from this equation are presented in Figure 2.

Consistent with our identification strategy, we find that exposure to female colleagues is unrelated to hiring decisions made prior to that exposure. Consistent with work by Dahl, Kotsadam & Rooth (2021) on the effect of exposure of mixed-gender teams on attitudes, we find no evidence for sustained or long-term effects of professional interactions on hiring decisions. This is perhaps unsurprising—service on judicial panels is a standard element of a judge's responsibilities, and judges are exposed to a new set of colleagues in each year. As a result, a year in which a judge interacted more than usual with female colleagues may temporarily affect behavior but not be memorable in the long term.

[13] Event study regressions omit judge fixed-effect because the positive relationship between exposure to female colleagues in period *t* and judge hiring decisions in period *t+1* biases coefficients downward for all hiring years other than year *t+1* in specifications that include judge fixed-effects. Event study regressions control for judge characteristics instead (see column 2 of Table 3).

## 6. Suggestive Evidence for Mechanisms

There are three broad pathways through which we might expect judges' exposure to female colleagues to affect their decision of who to hire. First, judges may change their beliefs about the suitability of women to the legal profession or their preferences for working closely with women after interacting with female colleagues. Second, they may experience social pressure from female colleagues to hire a diverse staff. Third, they may learn more about promising female law students from female colleagues. While all of these mechanisms result in greater opportunities for entry-level women, they may not be equally salient in other contexts. For example, if our results are primarily driven by direct encouragement to maintain a gender-diverse staff, we might expect smaller effects in contexts where hiring decisions are less connected to individuals or where colleagues are less outspoken than are federal judges, and larger effects in contexts where workers have less independence from their colleagues than do federal judges.

We examine the importance of the first two pathways by testing their empirical implications for the market for law clerks. If judges differ in their beliefs about the suitability of women to the practice of law or in their preferences for working with women, we would expect some judges to consistently hire more female clerks than others. Meanwhile, if judges preferred to maintain gender-diverse staffs, we would expect judges to hire a larger number of female clerks in years when a larger proportion of their current staff is male. We test these hypotheses using the following regression equation:

$$PctFem_{j,c,t+1} = \beta_1 AFF_{j,c,t} + \beta_2 \mathrm{CFF_{j,c,t}} + \delta X_{j,t} + \theta_{c,t} + \varepsilon_{j,c,t} \qquad (3)$$

$PctFem_{j,c,t+1}$ represents the fraction of clerks hired by judge *j*, in court *c*, in year *t+1* who are female. $AFF_{j,c,t}$ represents the average fraction of females on judge *j*'s staff in years other than year *t*. It is calculated as the number of female clerks hired by judge *j* in years other than year *t*

divided by the total number of clerks hired by judge *j* in years other than year *t*, and is used to reflect a judge's persistent preferences or beliefs influencing their hiring of female clerks. $CFF_{j,c,t}$ represents the current fraction of females on a judge's clerks at the time of hiring. It is calculated as the number of female clerks employed by judge *j* in year *t* divided by the total number of clerks employed by judge *j* in year *t*. $X_{j,t}$ is a set of judge characteristics and $\theta_{c,t}$ are court by year fixed effects. As shown in Table 5, the average gender composition of a judge's staff is positively correlated with their hiring of female clerks, but the current gender composition of their staff is uncorrelated with their hiring of female clerks. This suggests that while judges do differ in their likelihood of hiring female clerks, they are no more likely to hire women in years when they have few women in their staff.

We further investigate the second pathway--that judges care about maintaining a diverse staff—by examining the hiring decisions of judges who hire multiple clerks in the same year. Were judges indifferent to gender diversity of their staff, we would expect the gender of the second clerk hired to be unrelated to the gender of the first clerk hired, conditional on the judge's characteristics. As a result, we would expect the distribution of female hires by each judge to follow a Bernoulli distribution, where the probability of hiring a female clerk was given by the average fraction of females of each judge's hires ($AFF_{j,c,t}$). For instance, if 50% of a judge's clerks are female over the sample period, we would expect that one of two hires would be female 50% of the time, both would be female 25% of the time, and neither would be female 25% of the time. In contrast, if judges care about the gender diversity of their staffs, we would expect 1 of 2 hires to be female more than 50% of the time, and 0 or 2 hires to be female less than 25% of the time. We test this hypothesis in Appendix Table A.11. This compares the actual fraction of judges hiring 0, 1, 2, 3, 4, or 5 female clerks in a given year to predicted fractions. We obtain predicted fractions by

aggregating predicted values at the judge level given by Bernoulli distributions, where each judge's probability of hiring a female clerk is equals $AFF_{j,c,t}$. These two distributions are highly similar. In particular, the fraction of judges hiring no female clerks and hiring exclusively female clerks is similar to the fractions predicted by a Bernoulli distribution. We thus conclude that our results are unlikely to be driven by a desire to maintain a gender-diverse staff.

We also examine heterogeneity in the effect of interacting with female colleagues by both the characteristics of the influenced judges and the characteristics of the female colleagues with whom they interact. As shown in Appendix Table A.12, results are inconclusive because the effects are imprecisely estimated. However, one suggestive finding provides evidence against the third potential pathway—that judges learn about promising female clerks from female colleagues. Our point-estimates suggest that judges are no more likely to hire a woman after interacting with female colleagues with majority-female staffs than after interacting with female colleagues with majority-male staffs. This suggests that exposure to colleagues' networks is unlikely to drive our effects.

Together, while our data cannot definitively point to a single causal mechanism, it suggests that the effect of exposure to female colleagues on hiring decisions is most likely a consequence of changes to preferences or beliefs, rather than of an increased social cost of having a disproportionately male staff in a particular year or access to information about female job applicants.

## 7. Concluding Remarks

We find clear, causal evidence that contact with female colleagues increases the likelihood that judges hire women at the entry level of their profession. These findings are large, suggesting

that hiring a female judge to serve on a typical appellate court would lead the judge's male colleagues to hire an average of five additional female clerks over the next decade.[14]

This finding is relevant to discussions of diversification of leadership and affirmative action because it shows that ordinary workplace interactions can affect important economic decisions even among experienced decision-makers with established views. However, because this work is unique in its ability to estimate the effect of peer interactions among established professionals on their hiring practices, it is reasonable to wonder whether particular characteristics of the legal profession or the judiciary make it more or less susceptible to such effects.

One possibility is that peer exposure affects entry-level hiring more in contexts when new employees become mentees or proteges of their employers. Because appellate court clerkships are a major stepping stone to the American judiciary, judges care about the likelihood that a clerk will become a prominent lawyer or judge, and about how a clerk might conduct themselves as a judge in the future. As a result, a judge's beliefs about women's capabilities as judges may influence their hiring decisions even if those beliefs have no direct bearing on their performance as clerks (and thus are not primarily shaped by a judge's previous experience with female clerks). For example, expectations that women focus less on their careers after having children, or that women are less assertive, worse leaders, or less creative than men might all be more heavily influenced by working with female judges than by working with female clerks. Thus, diversity in the elite of a profession may affect entry-level hiring less in professions where mentorship and legacy play a smaller role.

[14] The average appellate circuit has 13 appointed judges, 25% of whom are women. Thus, hiring a woman as the 13th judge increases the fraction of interactions with a female judge by 1/12 for each of 9 male judges. 1/12*0.71*9*10=5.325

Another possibility is that interaction with female peers may be particularly salient in occupations where women's representation has increased during the career of the current leadership generation. Law is such a profession—while fewer than 5% of American law school graduates were women prior to 1970, 50% of recent law school cohorts are women.[15] Due to this generational change, the great majority of federal judges began their careers in a cohort composed predominantly of men, and at a time when virtually all professors and judges were men. It is thus reasonable to suspect that these judges have developed expectations about women in the profession that do not reflect the characteristics of the most recent cohort of lawyers. Thus, our results may be large precisely because judges have developed attitudes that can be counteracted with new evidence. Due to the dramatic increase in women's education and workforce engagement over the past seventy years, this generational change in the representation of women characterizes many high-earning professions.

A last possibility is that peer interaction has large effects on the legal profession because judges lack clear and objective criteria for hiring decisions, increasing the role of prior beliefs or prejudices (Coate & Loury 1993). While law students can signal quality by graduating from a top law schools or joining a legal review, the legal profession lacks highly objective measures of quality or productivity available in some engineering fields. As a result, peer interaction may matter less in fields where productivity and skill are more objectively measured.

## References

28 U.S. Code § 292. (1986). *Appointment, tenure, residence and salary of circuit judges, 292*. Retrieved from www.law.cornell.edu/uscode/text/28/44

Aigner, D., & Cain, G. (1977). Statistical Theories of Discrimination in Labor Markets. *Industrial and Labor Relations Review, 30*(2), 175-187. doi:10.1177/001979397703000204

Allport, G. (1954). *The Nature of Prejudice.* Cambridge, Mass: Perseus Books.

15 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. In *ISBN* (pp. 978–1–62722–880–0).

Ash, E., Chen, D., & Ornaghi, A. (2020). Stereotypes in High-Stakes Decisions: Evidence from U.S. Circuit Courts. In *The Warwick Economics Research Paper Series (TWERPS) 1256*. University of Warwick, Department of Economics.

Avery, C., Jolls, C., Posner, R. A., & Roth, A. E. (2007). The New Market for Federal Judicial Law Clerks. *University of Chicago Law Review, 74*, 447–486.

Bagues, M. F., & Esteve-Volart, B. (2010). Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment. *Review of Economic Studies, 77*, 1301–1328. doi:10.1111/j.1467-937x.2009.00601.x

Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R., & Topalova, P. (2009). Powerful Women: Does Exposure Reduce Bias. *Quarterly Journal of Economics, 124*, 1497–1540.

Bertrand, M., & Duflo, E. (2016). Field Experiments on Discrimination. *NBER*. doi:10.3386/w22014

Bertrand, M., Black, S. E., Jensen, S., & Lleras-Muney, A. (2019). Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway. *The Review of Economic Studies, 86*(1), 191-239. doi:10.1093/restud/rdy032

Blau, F., & Kahn, L. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature, 55*(3), 789-865. doi:10.1257/jel.20160995

Boisjoly, J., Duncan G., J., Kremer, M., Levy D., M., & Eccles, J. (2006). Empathy or Antipathy? The Impact of Diversity. *American Economic Review, 96*, 1890–1905.

Boyd C., L., Epstein, L., & Martin, A. D. (2010). Untangling the Causal Effects of Sex on Judging. *American Journal of Political Science, 54*, 389–411.

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

Chevrot-Bianco, E. (2021). It only takes a strong tie: Board gender quotas and network-based hiring. *IZA Workshop on Gender and Family Economics: "Women in Leadership" [Conference Paper]*. Retrieved from http://conference.iza.org/conference_files/Gender_2021/chevrot-bianco_e31446.pdf

Chilton A., S., & Levy, M. K. (2015). Challenging the Randomness of Panel Assignment in the Federal Courts of Appeals. In *Duke Law School Public Law and Legal Theory Series No. 2015-1; U of Chicago, Public Law Working Paper No.* doi:10.2139/ssrn.2520980

Coate, Stephen & Loury, Glenn (1993). Will Affirmative-Action Policies Eliminate Negative Stereotypes? *The American Economic Review*, 83(5), 1220-1240.

Courts, U. S. (2016). Judicial Business: Table B-12. *Judicial Business: Table B-12*. Retrieved from https://www.uscourts.gov/data-table-numbers/b-12

Dahl, G., Kotsadam, A., & Rooth, D.-O. (2021). Does Integration Change Gender Attitudes? The Effect of Randomly Assigning Women to Traditionally Male Teams. *The Quarterly Journal of Economics, 136*. doi:10.1093/qje/qjaa047

Disclosures, A. B. (2013). Enrollment and Degrees Awarded 1963 – 2012 Academic Years. *Enrollment and Degrees Awarded 1963 – 2012 Academic Years*.

Epstein, L., Martin, A. D., Segal, J. A., & Westerland, C. L. (2007). The Judicial Common Space. *Journal of Law, Economics, and Organization, 23*, 303–325. doi:10.1093/jleo/ewm024

Federal Judiciary, H. (2019). Web site of the Federal Judicial Center. *Web site of the Federal Judicial Center*. Washington, DC. Retrieved from http://www.fjc.gov.

Glynn A., N., & 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

Goldin, C., Kerr, S. P., Olivetti, C., & Barth, E. (2017). The Expanding Gender Earnings Gap: Evidence from the LEHD-2000 Census. *American Economic Review: Papers and Proceedings, 107*, 110-114.

Hess, A. G. (2015). The Collapse of the House that Ruth Built: The Impact of the Feeder System on Female Judges and the Federal Judiciary, 1970-2014. *Journal of Gender, Social Policy and the Law, 24*, 1–42.

Blau, F & Kahn, L (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature, 55*(3), 789-865. doi:10.1257/jel.20160995

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

Leagle. (2019). Retrieved from http://www.leagle.com/.

Levy, M. K. (2017). Panel Assignment in the Federal Courts of Appeals. *103 Cornell Law Review*, 65–116.

Levy, M. K. (2019). Visiting Judges. *California Law Review, 107*. Retrieved from https://ssrn.com/abstract=3328571

Lewis, J., Poole, K., Rosenthal, H., Boche, A., Rudkin, A., & Sonnet, L. (2019). Voteview: Congressional Roll-Call Votes Database. *Voteview: Congressional Roll-Call Votes Database*. Retrieved from https://voteview.com/

Lowe, M. (2021). Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration. *American Economic Review*, 1807-1844. doi:10.1257/aer.20191780

Maida, A., & Weber, A. (2020). Female Leadership and Gender Gap within Firms: Evidence from an Italian Board Reform. *ILR Review*, 001979392096199. doi:10.1177/0019793920961995

McCabe, P. G. (2014). *A Guide to the Federal Magistrate Judges System.* Technical report,, Federal Bar Association.

Mousa, S. (2020). Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq. *Science*, 866-870. doi:10.1126/science.abb3153

N.A.L.P. (2019). Clerkship Study Law Student Findings. *Clerkship Study Law Student Findings*.

Paluck, E. L., Green, S. A., & Green, D. (2018). The Contact Hypothesis Re-evaluated. *Behavioural Public Policy, 3*, 129–158. doi:10.1017/bpp.2018.25.

Posner, R., Roth, A. E., Avery, C., & Jolls, C. (2001). The Market for Federal Judicial Law Clerks. *University of Chicago Law Review, 68*, 793–902.

Rhinehart, L. K. (1994). Is There Gender Bias in the Judicial Law Clerk Selection Process. *Georgetown Law Journal*, 575-603.

Stearns M., L., & Abramowicz, M. (2005). Defining Dicta. *Faculty Scholarship, 119*. Retrieved from https://digitalcommons.law.umaryland.edu/fac_pubs/119

Washington E., L. (2008). Female Socialization: How Daughters Affect Their Legislator Fathers. *American Economic Review, 98*, 311–32.

# Tables and Figures

## Tables

Table 1: Characteristics of Judges and Clerks in Sample

| | Full Sample Mean (SD) | Male Mean (SD) | Female Mean (SD) | Full Sample Min / Max | Sample Size Male / Female | M/F Diff |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Judge characteristics | | | | | | |
| Female | 0.2562 (0.4367) | 0.0000 (0.0000) | 1.0000 (0.0000) | 0 / 1 | 996 / 343 | *** |
| Asian | 0.0127 (0.1120) | 0.0151 (0.1219) | 0.0058 (0.0762) | 0 / 1 | 996 / 343 | |
| Black | 0.0941 (0.2921) | 0.0894 (0.2854) | 0.1079 (0.3107) | 0 / 1 | 996 / 343 | |
| Hispanic | 0.0635 (0.2439) | 0.0653 (0.2471) | 0.0583 (0.2347) | 0 / 1 | 996 / 343 | |
| Age (decades) | 6.4277 (1.0089) | 6.5443 (1.0186) | 6.0892 (0.8996) | 3.6 / 9.2 | 996 / 343 | *** |
| Decades on current court | 1.5999 (0.9882) | 1.7105 (0.9849) | 1.2787 (0.9269) | 0 / 4.4 | 996 / 343 | *** |
| Ideology score | 0.0681 (0.3614) | 0.0888 (0.3550) | 0.0081 (0.3736) | -0.521 / 0.693 | 996 / 343 | *** |
| Republican | 0.5400 (0.4986) | 0.5763 (0.4944) | 0.4344 (0.4964) | 0 / 1 | 996 / 343 | *** |
| Number of clerks hired in year | 2.8656 (1.2262) | 2.8072 (1.2167) | 3.0350 (1.2397) | 1 / 7 | 996 / 343 | *** |
| Number of cases heard in year | 57.7371 (40.4221) | 58.1908 (41.1069) | 56.4198 (38.3920) | 1 / 208 | 996 / 343 | |
| At least one female clerk hired | 0.6923 (0.4617) | 0.6807 (0.4664) | 0.7259 (0.4467) | 0 / 1 | 996 / 343 | |
| Percent of clerks that are female | 0.4205 (0.4937) | 0.4210 (0.4938) | 0.4192 (0.4936) | 0.00 / 1.00 | 3637 / 1274 | |

*Notes:* This table presents the average values of judge-level covariates in the analysis sample. For all variables other than percent of clerks that are female, the sample is at the judge by year level. For Percent of clerks that are female, the sample is at the clerk by year level. *Source*: *Judicial Yellow Books*, case dataset collected by authors.

Table 2: Balance Tests for Random Assignment to Panels

| Dependent variable: Fraction of co-panelists who are female | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Age | 0.0019 | 0.0015 | 0.0002 | 0.0054 |
| | (0.0045) | (0.0029) | (0.0035) | (0.0056) |
| Asian | -0.0136 | -0.034 | -0.037 | -0.0002 |
| | (0.0315) | (0.0313) | (0.0302) | (0.0198) |
| Black | 0.0044 | -0.0028 | 0.0052 | 0.01 |
| | (0.0193) | (0.0074) | (0.0088) | (0.0145) |
| Hispanic | 0.0148 | 0.0008 | -0.0161 | 0.0163 |
| | (0.0155) | (0.0114) | (0.0103) | (0.0317) |
| Years on current court | 0.0106** | -0.0004 | -0.0004 | 0.0005 |
| | (0.0046) | (0.0027) | (0.0031) | (0.0049) |
| Ideology score | 0.0244* | -0.0017 | -0.0039 | 0.0117 |
| | (0.014) | (0.0088) | (0.0106) | (0.0198) |
| Republican | 0.0112 | 0.0018 | 0.0003 | 0.008 |
| | (0.0096) | (0.0057) | (0.0066) | (0.0132) |
| Percent of current staff female | -0.0017 | 0.0157 | 0.0204* | 0.0161 |
| | (0.0139) | (0.0099) | (0.012) | (0.022) |
| Court by year fixed effects | No | Yes | Yes | Yes |
| Sample | All | All | Male | Female |
| F-Stat (P-val) | 1.4215 | 0.7416 | 1.0727 | 0.9399 |
| | (0.1894) | (0.6547) | (0.385) | (0.4928) |
| Observations | 1339 | 1339 | 996 | 343 |

*Notes:* This 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. It 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). 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.

Table 3: Effect of Serving with Female Judges on Hiring Decisions

| Dep Var: Probability of hiring any female clerk in next year | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Fraction of co-panelists who are female | 0.5901*** (0.1948) | 0.6590*** (0.1863) | 0.7628*** (0.2623) | 0.7120*** (0.2603) | 0.7258*** (0.2726) |
| Number of clerks hired | | | | 0.0724*** (0.0189) | 0.0701*** (0.0200) |
| Percent of current staff female | | | | | -0.0725 (0.0745) |
| Court by year fixed effects | Yes | Yes | Yes | Yes | Yes |
| Judge Characteristics | | Yes | | | |
| Judge Fixed Effects | | | Yes | Yes | Yes |
| Observations | 996 | 996 | 996 | 996 | 925 |
| Dependent variable mean | 0.68072 | 0.68072 | 0.68072 | 0.68072 | 0.68072 |

*Notes:* This table reports OLS estimation results from the regressions described in equation 1 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 year *t+1*, 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. Judge characteristics include quadratics of judge age, experience in current position and ideology, as well as judge gender, race and party of nominating president. Significance levels are: * 10%, ** 5%, *** 1%. *Source: Judicial Yellow Books*, case dataset collected by authors.

Table 4: 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 | -0.0495 | 0.0238 | 0.0306 | -0.1514 | -0.0935 | 0.1090 |
| | (0.1102) | (0.1104) | (0.1530) | (0.1233) | (0.1239) | (0.1548) |
| Number of clerks hired | | | 0.0776*** | | | 0.0744*** |
| | | | (0.0183) | | | (0.0189) |
| Court by year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Judge Characteristics | | Yes | | | Yes | |
| Judge fixed effects | | | Yes | | | Yes |
| Observations | 1033 | 1029 | 1029 | 972 | 972 | 972 |
| Dependent variable mean | 0.6873 | 0.6881 | 0.6881 | 0.6811 | 0.6811 | 0.6811 |

*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*, conditional on hiring any clerk. 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*, 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. Judge characteristics include quadratics of judge age, experience in current position and ideology, as well as judge gender, race and party of nominating president. Significance levels are: * 10%, ** 5%, *** 1%. *Source: Judicial Yellow Books*, case dataset collected by authors.

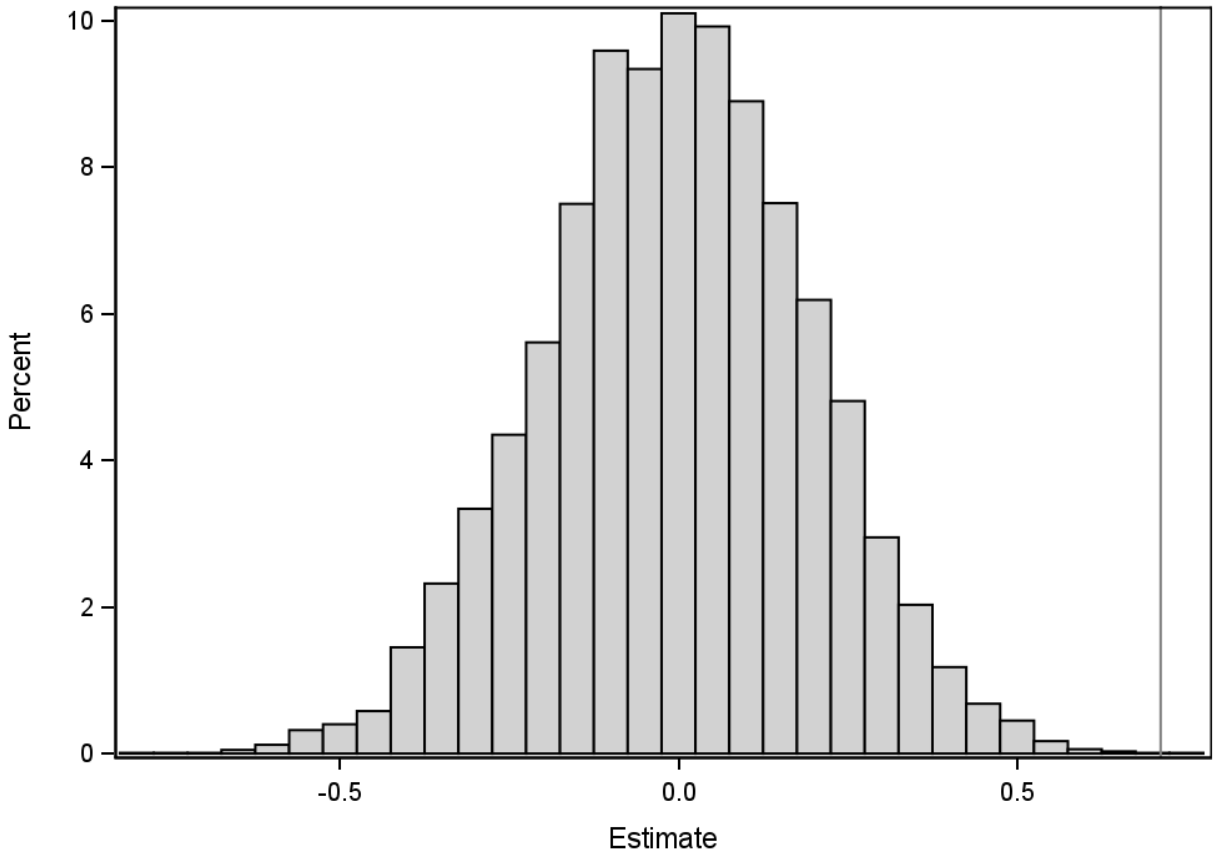
Table 5: Persistent Preferences and Staff Gender Composition on Female Hiring

| Dep Var: Fraction of hires in next year who are female | (1) | (2) | (3) |
|---|---|---|---|
| Persistent hiring rate | 0.3390*** | 0.3065*** | 0.2678*** |
| | (0.0942) | (0.0960) | (0.1001) |
| Percent of current staff female | 0.0217 | 0.0179 | 0.0342 |
| | (0.0491) | (0.0492) | (0.0489) |
| Asian | | -0.0022 | -0.0067 |
| | | (0.0869) | (0.0892) |
| Black | | 0.0409 | 0.0181 |
| | | (0.0461) | (0.0506) |
| Hispanic | | 0.0810* | 0.0507 |
| | | (0.0428) | (0.0462) |
| Age | | -0.0030 | 0.0241 |
| | | (0.0052) | (0.0448) |
| $Age^2$ | | 0.0000 | -0.0007 |
| | | (0.0001) | (0.0009) |
| Years on current court | | | 0.1326 |
| | | | (0.1173) |
| Years on current $court^2$ | | | -0.2118 |
| | | | (0.4056) |
| Ideology score | | | -0.0959 |
| | | | (0.0713) |
| Ideology $score^2$ | | | -0.3945** |
| | | | (0.1821) |
| Republican | | | 0.0893 |
| | | | (0.1062) |
| Court by year fixed effects | Yes | Yes | Yes |
| Observations | 939 | 939 | 939 |
| Dependent variable mean | 0.4020 | 0.4020 | 0.4020 |

Notes: This table reports OLS estimation results from a regression of the fraction of hires in year t who are female on the fraction of all hires in years other than year t who are female and on the fraction of staff in year t who are female, as described in equation 3 in the text. Note that these regressions do not include judge fixed-effects. Standard errors are robust and clustered at the judge level. The dependent variable is the fraction of clerks hired in the following year who are female, conditional on hiring any clerk. Significance levels are: * 10%, ** 5%, *** 1%. *Source: Judicial Yellow Books*, case dataset collected by authors.

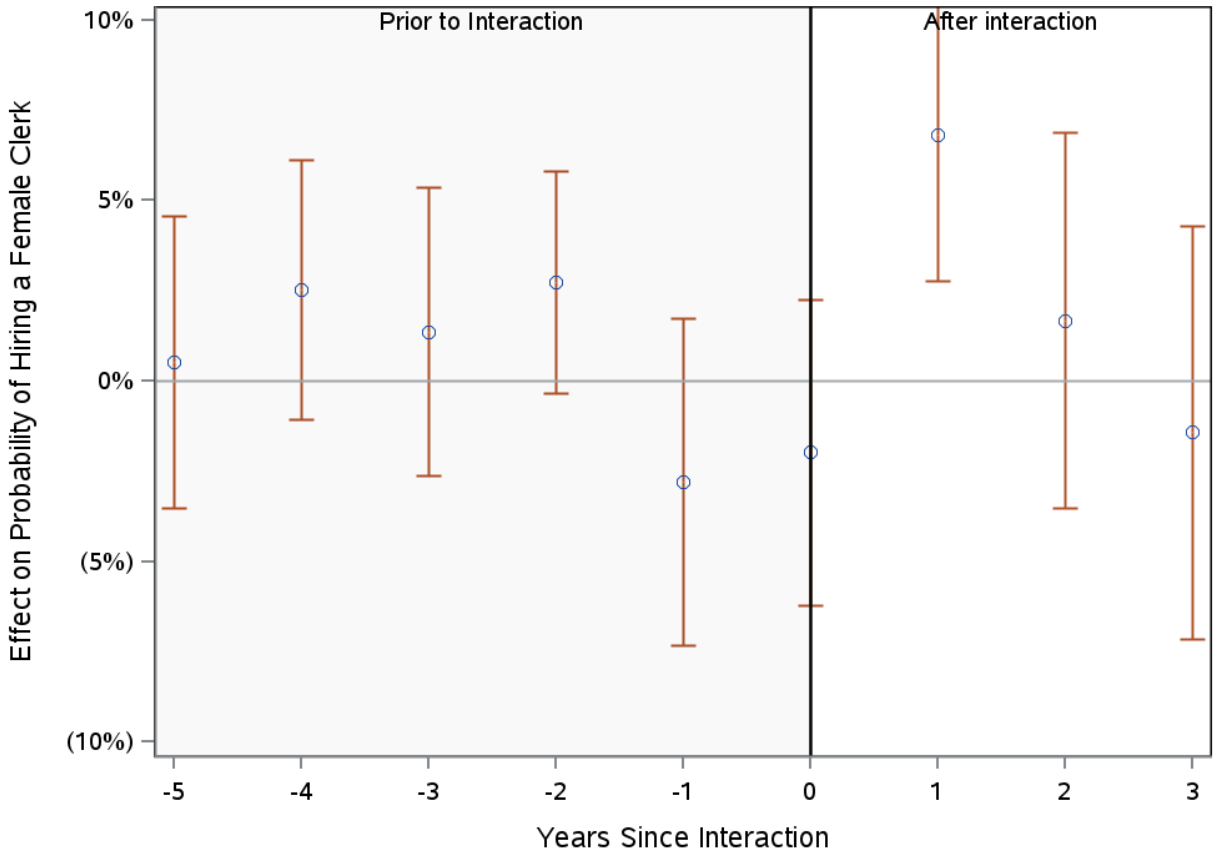
## Figures

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




*Notes:* The figure shows distribution of coefficients obtained from the OLS specification shown in column 4 of table 3 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. We performed 10,000 iterations of this resampling procedure. The vertical line represents the actual estimate obtained in column 4 Table 3.

Figure 2: Event Study: Effect of Fraction Female Copanelists in Year *t* on Probability of Hiring at Least One Female Clerk in Year *t+k*




*Notes:* This figure reports the effect of a one standard-deviation increase in interactions with female judges in year *t* on the probability of hiring at least one female clerk in year *t+k*, estimated using equation 2, with 95% confidence intervals. Standard errors are robust and clustered at the judge level. Controls include quadratics of judge age, experience in current position and ideology, as well as judge gender, race, and party of nominating president. Regressions do not control for judge fixed effects. *Source*: *Judicial Yellow Books*, case dataset collected by authors.